IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Target-Dependent Sentiment Classification with BERT

ZHENGJIE GAO, AO FENG, XINYU SONG AND XI WU
Department of Computer Science, Chengdu University of Information Technology, Chengdu 610225, China

Corresponding author: Ao Feng (e-mail: abraham.feng@gmail.com).

This work was supported in part by the Research Innovation Team Fund (Award No. 18TD0026) and Youth Technology Fund (Award No. 2017JQ0030) from the Department of Education, and in part by the Seedling Project of Science and Technology Innovation (Project No. 2018115) from the Science Technology Department, Sichuan Province.

ABSTRACT Research on machine assisted text analysis follows the rapid development of digital media, and sentiment analysis is among the prevalent applications. Traditional sentiment analysis methods require complex feature engineering, and embedding representations have dominated leaderboards for a long time. However, the context-independent nature limits their representative power in rich context, hurting performance in Natural Language Processing (NLP) tasks. Bidirectional Encoder Representations from Transformers (BERT), among other pre-trained language models, beats existing best results in eleven NLP tasks (including sentence-level sentiment classification) by a large margin, which makes it the new baseline of text representation. As a more challenging task, fewer applications of BERT have been observed for sentiment classification at the aspect level. We implement three target-dependent variations of the $\mathbf{BERT}_{\mathbf{base}}$ model, with positioned output at the target terms and an optional sentence with the target built in. Experiments on three data collections show that our TD-BERT model achieves new state-of-the-art performance, in comparison to traditional feature engineering methods, embedding-based models and earlier applications of BERT. With the successful application of BERT in many NLP tasks, our experiments try to verify if its context-aware representation can achieve similar performance improvement in aspect-based sentiment analysis. Surprisingly, coupling it with complex neural networks that used to work well with embedding representations does not show much value, sometimes with performance below the vanilla BERT-FC implementation. On the other hand, incorporation of target information shows stable accuracy improvement, and the most effective way of utilizing that information is displayed through the experiment.

INDEX TERMS Deep Learning, Neural Networks, Sentiment Analysis, BERT

I. INTRODUCTION

The size of digital media is growing at an exploding speed, which makes information consumption a challenging task. A large portion of the digital media is user generated, but manually locating the required information is beyond the ability of any human being. Machine assisted media processing is valuable for many recipients, including governments, companies and individuals, while its applications include stock price prediction, product recommendation, opinion poll, etc. All these require accurate extraction of main entities, together with opinions or attitudes expressed by the author.

Sentiment analysis is a fundamental task in Natural Language Processing (NLP). It is crucial for understanding user generated text in news reports, product reviews, or social discussions. Aspect-Based Sentiment Analysis (ABSA) [12, 19, 24, 28] is a fine-grained task in sentiment analysis, which aims to identify the sentiment polarity (e.g., positive, negative, neutral, conflict) of an aspect category [28] or a target (also called an aspect term [28]). In this paper, we focus on target-dependent sentiment classification [6, 15, 34, 38, 41]. As specific instances of aspects, targets explicitly occur in sentences, and the polarity of sentiment towards them needs to be identified separately. As illustrated in figure 1, in a sentence "I bought a mobile phone, its camera is wonderful but battery life is short", the sentiment polarity for term "camera" is positive, for "battery" it is negative, and for "mobile phone" conflicted sentiments are found in the same sentence, as both positive (wonderful) and negative (life is short) sentiments are expressed towards the same target. The target information is important in its corresponding sentiment, as one sentence can refer to many targets, each with its own context. It is hard to determine the sentiment for a target term without accurate

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2946594, IEEE Access

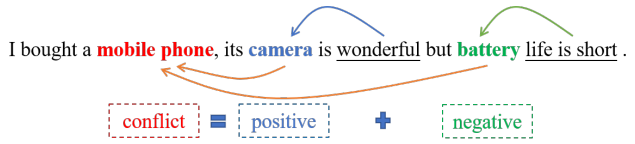


FIGURE 1. An example of consumer review with three targets which have different sentiment polarities. The targets are highlighted with different colors, the opinions are underlined and point to their corresponding targets.

aspect information, which accounts for a large portion of sentiment classification errors [15].

Traditional target-dependent sentiment classification focuses on feature engineering to get the most out of a classifier (e.g., Support Vector Machines) [15, 18, 42]. Such methods need laborious feature engineering work and/or massive linguistic resources, which are time-consuming, error-prone and require extensive domain knowledge from *experts*. Multiple sentiment lexicons are built for this purpose [22, 30, 37], taking a large amount of human labor, but they are difficult to be transferred into another domain.

With the recent development of deep learning, a large number of neural network models are present in this NLP task [4, 6, 13, 20, 34, 36, 38, 39, 44]. With neural networks' capacity of learning representation from data without complex feature engineering, deep learning becomes the hottest research model in this area as well as many others. The mainstream neural networks are Long Short-Term Memory (LSTM) [9] and memory networks [35], with attention mechanism [1] frequently used to locate the right context. Recursive neural networks [6, 23], gated neural networks [47, 49] and convolutional neural networks [13] are less popular.

Unlike the computer vision domain in which transfer learning is a common approach in low-level feature extraction, NLP tasks almost always restart training from scratch in each application. The earlier exception may be the use of text embeddings [21, 26] that are trained on large-scale unlabeled text collections, but they suffer from the context-independence assumption that each term has the same embedding despite its surrounding context. More recently, pre-trained language models such as ULMFiT [11], OpenAI GPT [31], ELMo [27] and BERT [5] have shown great power in the semantic expressiveness of text. Among them, Bidirectional Encoder Representations from Transformers (BERT) has achieve excellent results in sentence-level sentiment classification (SST-2), together with ten other tasks. Despite its burgeoning popularity in many NLP applications, wide application of BERT has not been observed in ABSA.

In this paper, we investigate related work in feature engineering models, embedding-based neural networks, and also try BERT with traditional networks. Then we propose a Target-Dependent BERT (TD-BERT) model with several variations. On SemEval-2014 and a Twitter dataset, we compare the classification accuracy of aspect term polarity (SemEval-2014 Task 4 subtask 2) of these methods, and our TD-BERT models consistently outperform others, including recent BERT-based models. The experiments show that complex neural networks that used to return good results with embeddings do not fit well with BERT, while incorporation of target information into BERT yields stable performance boost.

Main contributions of this article include: 1. We utilize BERT in aspect-level sentiment classification, and achieve new state-of-the-art performance on three public datasets. 2. Swapping embedding representations with BERT does not naturally improve the performance of existing neural network models, as they are better tuned with the context-independent representation. 3. Incorporation of target information is a key factor in BERT's performance improvement, and we show several simple but effective strategies to implement that.

II. RELATED WORK

As shown in [48], there are three important tasks in ABSA. The first task is to represent the whole context that a target appears in, the second generates a representation of the target itself, and the last task is to identify the important part of context for the sentiment judgment of the specific target. In any natural language processing task, representation of the text (including the target and its context) is a key issue.

With proper representation of target and context information, the next phase designs a classification model and generates the sentiment label for a target. There are more choices in this stage, from traditional machine learning models to various types of neural networks. Their performance greatly rely on the expressiveness power of text representation from the previous step, but the classifier itself is also of great importance.

A. TEXT REPRESENTATION

In traditional information retrieval, a term is represented by a one-hot vector, in which one dimension is set to 1 which corresponds to the index of the term while all other dimensions are zero. Such representation is straight-forward, but suffers from its high dimensionality and poor correlation among similar terms. pLSI [10] and LDA [3] relax the independence assumption by introducing an *aspect* or *topic*, but the mapping does not simplify the original vector which is still in vocabulary size V.

Starting with [2], the one-hot vector is replaced by a low-dimensional distributed representation. After that, word embedding becomes the standard technique for obtaining pretrained vector representations from large unlabeled corpora. Word2Vec [21] and Global Vectors (GloVe) [26] are the most popular embedding representations to capture syntactic and semantic features of text. A large number of experiments have demonstrated that pretrained word embeddings can improve performance on a variety of NLP tasks [17, 29]. However, existing word embedding methods, which use limited window size, cannot exploit semantic information in the global context. In addition, such an algorithm transforms a

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2946594, IEEE Access

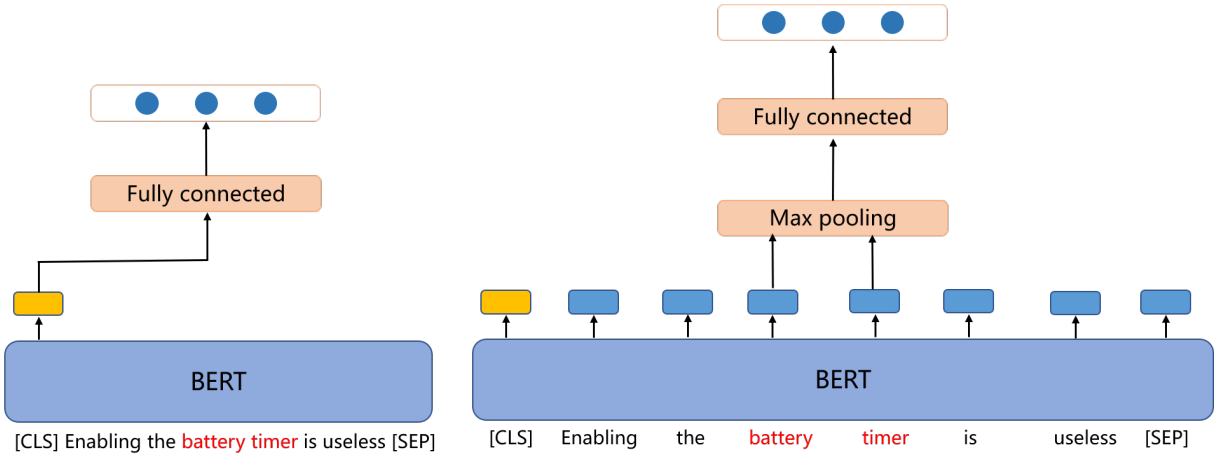


FIGURE 2. The architecture of BERT-FC (left) and TD-BERT (right)

word into a stable vector. As a result, the vector is unable to accurately represent its context at different locations.

Recently, language representation models with context, like ULMFiT [11], OpenAI GPT [31], ELMo [27] and BERT[5], are designed by jointly conditioning on both left and right context with deep neural networks. Furthermore, these models can dynamically adjust the word vector according to its context. Our work mainly relies on BERT, which provides us with a strong baseline, and we further modify its standard network structure to incorporate the target information.

B. CONVENTIONAL NEURAL NETWORKS

In order to explicitly distinguish target words from the context, Vo and Zhang [41] divide the original sentence into three parts according to a given target, a left context, a right context and the target itself. Tang et al. [38] propose target-dependent LSTM (TD-LSTM) to capture the aspect information when modeling sentences. A forward LSTM and a backward LSTM towards target words are used to capture the information before and after the aspect. Wang et al. [44] adopt attention mechanism to concentrate on corresponding parts of a sentence when different aspects are taken as input.

Previous approaches have revealed the importance of targets in ABSA and developed various methods with the goal of precisely modeling their contexts via generating target-specific representations. However, these studies usually ignore the separate modeling of targets. Ma et al. [20] argue that both the target and its context deserve special treatment and their own representations need to be trained via interactive learning. They propose an Interactive Attention Networks (IAN) to learn the representations for target and context separately.

Convolutional Neural Networks (CNN), Recursive Neural Networks [6] and Gated Neural Networks [49] are used less frequently in aspect-level sentiment classification. CNN has been successfully applied in sentence-level sentiment classification [16, 17, 50]. Huang and Carley [13] are claimed to be the first to use CNN for sentiment classification at the aspect level, in which they incorporate aspect information into CNN via parameterized filters and parameterized gates, resulting in good performance metrics on SemEval-2014 datasets.

C. MEMORY NETWORK

Tang et al. [39] develop a deep memory network for aspect-level sentiment classification, which is inspired by the success of computational models with attention mechanism and explicit memory [1, 7, 35]. They employ an attention mechanism with external memory to capture the importance of each context word with respect to the given target aspect. This approach explicitly captures the importance of each context word when inferring the sentiment polarity of the aspect. The importance degree and text representation are calculated with multiple computational layers, each establishing a neural attention model over an external memory. Chen et al. [4] propose a recurrent attention network to better capture the sentiment of complicated contexts. To achieve that, their proposed model uses a recurrent/dynamic attention structure and learns a non-linear combination of the attention in GRUs. Zhu and Qian [51] show a novel deep memory network with extra memory that can utilize the information of aspects and terms at the same time. The main memory is used to capture the important context words for sentiment classification. In addition, an auxiliary memory is built to implicitly convert aspects and terms into each other, and then they are both fed into the main memory. With the interaction between two memory blocks, the features of aspects and terms can be learned simultaneously.

D. BERT-BASED NETWORKS

Most of the previous approaches model the relation between target words and their context with LSTM and attention. However, LSTMs are difficult to parallelize and truncated back-propagation through time brings difficulty in remem-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2946594, IEEE Access

bering long-term patterns. To address this issue, Song et al. [34] propose an Attentional Encoder Network (AEN) without a recurrent structure and employ attention based encoders for the modeling between context and target. Sun et al. [36] construct an auxiliary sentence from the aspect and convert aspect-based sentiment classification into a sentence-pair classification task. Xu et al. [46] explore the potential of turning customer reviews into a large source of knowledge, which can be exploited to answer user questions. The new task is named Review Reading Comprehension (RRC). They explore a novel post-training approach to enhance the performance by fine-tuning the BERT network for RRC. Then Aspect Sentiment Classification (ASC) is converted into a special Machine Reading Comprehension (MRC) problem [32, 33], in which all questions are about the polarity of a given aspect.

III. OUR APPROACH

In this section, we introduce our method for target-dependent sentiment classification, which is based on BERT. We first describe BERT for a general sentiment classification task, then introduce our base model Target-Dependent BERT (TD-BERT) and another two variants that combine sentence-pair classification with TD-BERT.

A. PROBLEM DEFINITION AND NOTATIONS

A target-dependent sentiment classification task usually predicts the sentiment polarity of a tuple (s, t) which consisting of a sentence and a target. The sentence

$$s = [w_1, w_2, \ldots, w_i, \ldots, w_n] \tag{1}$$

consists of n words, and the target

$$t = [w_i, w_{i+1}, \ldots, w_{i+m-1}] \tag{2}$$

contains m words, while t is a subsequence of s. The goal of this task is to determine the sentiment polarity y of sentence s towards the target t, where

$$\mathrm{y} \in \{\mathrm{positive}, \mathrm{negative}, \mathrm{neutral}, \mathrm{conflict}\} \tag{3}$$

For example, the sentence

$$s1 = \text{"great food but the service was dreadful!"} \tag{4}$$

is *positive* for "food" and *negative* for "service". In another example, the sentence

$$\mathrm{s2} = \text{"The sound as mentioned earlier isn't the best, but it can be solved with headphones."} \tag{5}$$

is *conflict* for "sound" as there are multiple cases of sentiment expression towards different polarity, and *neutral* for "headphones".

B. BERT

BERT [5] is a new language representation model, which uses a bidirectional Transformer [40] network to pre-train a language model on a large corpus, and fine-tunes the pre-trained model on other tasks. The task-specific BERT design is able to represent either a single sentence or a pair of sentences as a consecutive array of tokens. For a given token, its input representation is constructed by summing its corresponding token, segment, and position embeddings. For a classification task, the first word of the sequence is identified with a unique token [CLS], and a fully-connected layer is connected at the [CLS] position of the last encoder layer, finally a softmax layer completes the sentence or sentence-pair classification.

BERT has two parameter intensive settings:

$\mathbf{BERT_{base}}$: The number of Transformer blocks is 12, the hidden layer size is 768, the number of self-attention heads is 12, and the total number of parameters for the pre-trained model is 110M.

$\mathbf{BERT_{large}}$: The number of Transformer blocks is 24, the hidden layer size is 1024, the number of self-attention heads is 16, and the total number of parameters for the pre-trained model is 340M.

The $\mathbf{BERT_{large}}$ model requires significantly more memory than $\mathbf{BERT_{base}}$. As a result, the max batch size for $\mathbf{BERT_{large}}$ is so small on a normal GPU with 12GB of RAM that it actually hurts the model accuracy, regardless of the learning rate [5]. Therefore, we used $\mathbf{BERT_{base}}$ as our base model for further processing.

C. TARGET-DEPENDENT BERT

When dealing with sentence-level sentiment classification, the output at the [CLS] tag of BERT is directly followed by a fully-connected layer for classification, which we called BERT-FC. We can see that it does not incorporate any target information in its classification input. In Figure 2, BERT-FC is on the left side, and the right represents the architecture of Target-Dependent BERT (TD-BERT), which takes output from the target terms (in red). When there are multiple target terms, a max-pooling operation is taken before data is fed to the next fully-connected layer. As shown in Figure 2, the main difference between TD-BERT and BERT-FC is that TD-BERT takes the positioned output at the target words as input for classification instead of the first [CLS] tag.

BERT uses WordPiece [45] as its tokenizer. After the multi-layer bidirectional Transformer network, the word vector matrix Sr of the sentence s is represented by the hidden status of the last layer.

$$\mathrm{Sr} = [x_0, x_1, x_2, \ldots, x_i, \ldots, x_n, x_{n+1}] \tag{6}$$

$Sr \in R^{(n+2)\times d}$, where d is the dimension of hidden status, n is the length of the sentence. x_0 is the vector of the sentence classification mark [CLS], and x_{n+1} is the vector of the sentence separator or end [SEP].

The target words are represented by a sub-matrix of Sr

$$\mathrm{Tr} = [x_i, x_{i+1}, \ldots, x_{i+m-1}] \tag{7}$$

$Tr \in R^{m\times d}$, and m represents the length of the target. The max-pooling operation is applied to the target vectors, as the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2946594, IEEE Access

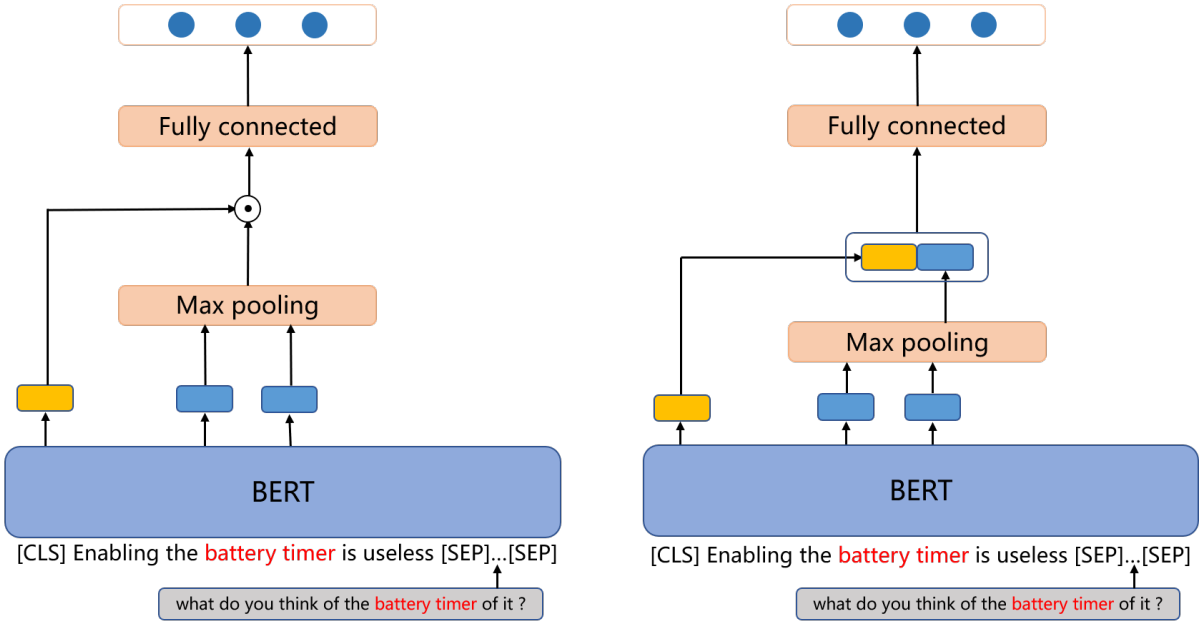


FIGURE 3. The architecture of TD-BERT-QA-MUL (left) and TD-BERT-QA-CON (right)

TABLE 1. Illustration of hard data with target words in bold

Sentence	Target	Polarity	Hard data	Remark
Nevertheless the **food** itself is pretty good.	food	positive	No	Only one target
A few tips: skip the **turnip cake**, **roast pork buns** and **egg custards**	turnip cake	negative	No	All targets have the same polarity
	roast pork buns	negative		
	egg custards	negative		
Fan vents to the side, so no **cooling pad** needed, great **feature!**	Fan	positive	Yes	Multiple targets with different polarities
	cooling pad	neutral		
	feature	positive		

most important features at each dimension is selected from all words in the target.

$$\mathrm{V} = \max\{\mathrm{Tr}, \mathrm{dim} = 0\}, \mathrm{V} \in R^{1\times d} \tag{8}$$

Finally, V is fed into a fully-connected layer and softmax for classification.

D. TWO VARIANTS

Sun et al. [36] construct an auxiliary sentence from the aspect and convert ABSA into a sentence-pair classification task. Given a sentence s and a target t, we can generate an auxiliary question "What do you think of the t of it?" similarly. Combining the idea with TD-BERT introduces many possibilities, and we try both element-wise multiplication and concatenation for features extracted from two base models. One of the base models is TD-BERT described above, the other is equivalent to [36], which adds an auxiliary sentence to the end of the original text and outputs data from the [CLS] location. The two variants are shown in Figure 3.

TD-BERT-QA-MUL: After normalizing the feature information of target words $\mathrm{V} \in R^{1\times d}$, take its element-wise product with the sentence pair output $\mathrm{P} \in R^{1\times d}$. The product is fed into a fully-connected layer.

TD-BERT-QA-CON: The target information V and the result P of sentence-pair classification are concatenated as a vector $d \in R^{1\times 2d}$ before going into the fully-connected layer. As the input dimension is larger, this layer has twice the parameters as in TD-BERT-QA-MUL.

IV. EXPERIMENTS

A. DATASETS

There are three datasets in the experiment, as shown in Table 2. The first two are from SemEval-2014 task 4[1] [28], including data in the restaurant and laptop domains, which are widely used in previous work. The last one is a collection of tweets collected by [6]. It is worth noting that the fourth classification category exists in the first two datasets - conflict, which means that a sentence expresses both positive and negative opinions towards an aspect or target. For example, "Certainly not the best sushi in New York, however, it is always fresh" [28]. Some existing work [4, 13, 34, 39, 44, 46] remove "conflict" samples from the data since the number of such instances is very small, while keeping them in the training data makes the dataset extremely unbalanced. With

[1]http://alt.qcri.org/semeval2014/task4/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2946594, IEEE Access

TABLE 2. Statistics of the experiment datasets

Dataset	Positive	Negative	Neutral	Conflict	Total
Laptop-Train	987	866	460	45	2358
Laptop-Test	341	128	169	16	654
Laptop-Train-Hard	159	147	173	17	496
Laptop-Test-Hard	31	25	49	3	108
Restaurant-Train	2164	805	633	91	3693
Restaurant-Test	728	196	196	14	1134
Restaurant-Train-Hard	379	323	293	43	1038
Restaurant-Test-Hard	92	62	83	8	245
Twitter-Train	1561	1560	3127	-	6248
Twitter-Test	173	173	346	-	692

such a revision, their results are not directly comparable to those from the SemEval-2014 [28] evaluation.

Statistics of *hard* data [47] are also shown in Table 2, with some instances displayed in Table 1. Here *hard* means that the sentence has multiple aspect labels associated with different sentiment polarities. Without them, a sentence-level sentiment classifier may be good enough for a single-target sentence (or all sentiment labels are consistent for multiple targets). Note that a sentence contains only one target in the Twitter dataset [43].

B. EXPERIMENT SETTINGS

All models are implemented in PyTorch [25], and the pre-trained uncased base model of BERT [2] is fine-tuned on a single NVIDIA RTX 2080Ti GPU (12GB RAM). Hyperparameters in the experiment are shown in Table 3. Except for a different max epoch number, we keep the recommended parameters from the official BERT code [3].

TABLE 3. Hyperparameters used in the experiment

Parameter	Value
Dropout rate	0.1
Batch size	32
Learning rate	2e-5
Max epoch	6
Max sequence length	128
Optimizer	Adam

C. RESULTS

We use the classification accuracy metric to measure the performance of our model and previous systems. To show the effectiveness of our model, we compare it to many baseline methods, as listed below:

DCU & NRC-Canada: Both DCU [42] and NRC-Canada [18] rely on an SVM classifier with features mainly from n-grams, parse trees, and several out-of-domain, publicly available sentiment lexicons (e.g., MPQA, SentiWordnet and Bing Liu's Opinion Lexicon). DCU and NRC-Canada ranked the top two at SemEval-2014 task 4 subtask 2 (aspect term polarities) [28].

Rec-NN [6] first applies rules to transform the dependency tree of a sentence and puts the target at the tree root, and then learns the semantic composition of the sentence via Recursive Neural Networks for classification.

TD-LSTM [38] designs two LSTM networks to model the context before and after the target, one in the left-to-right direction and the other in the reverse order. Then the last hidden states of the two networks are concatenated for predicting the sentiment polarity of the target.

ATAE-LSTM [44] attaches the target embedding onto the representation of each word, and LSTM with attention mechanism is applied to form the final representation for classification.

MemNet [39] is an end-to-end deep memory network that uses multiple computational layers to capture the importance of each context word.

RAM [4] adopts a multiple-attention mechanism to capture sentiment features separated by a long distance. A recurrent neural network is used to combine multiple attention outputs and strengthen the expressive power of MemNet.

IAN [20] uses two LSTM networks to model the sentence and target terms respectively. Then the target's hidden states and the context's hidden states are placed in parallel to supervise the generation of attention vectors interactively. Finally, it generates a sentence representation and a target representation based on these attention vectors.

GCAE [47] is a convolutional neural network with gating mechanism. The Gated Tanh-ReLU Units can selectively output the sentiment features according to the given aspect.

AOA [14] utilizes an Attention-over-Attention module to capture the interaction between aspects and context sentences in a joint way. With this design, AOA can learn the important parts in the aspects and context sentences, which generates the final representation of the sentence.

BERT-FC is a pre-trained BERT model with a fully-connected layer and softmax for classification, as shown in Figure 2. This method does not consider any target information, so it always returns the same sentiment polarity no matter which target is selected. It represent the basic implementation of BERT, allowing other models to take advantage of their awareness of the target information.

BERT-pair-QA-M [36] constructs an auxiliary question from the given aspect term and fine-tunes the pre-trained model from BERT for sentence-pair classification. [36] and

[2]https://storage.googleapis.com/bert_models/2018_10_18/uncased_L-12_H-768_A-12.zip

[3]https://github.com/google-research/bert

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2946594, IEEE Access

TABLE 4. Performance comparison with classification accuracy and F1 value on the test set as the evaluation metrics. 3way stands for 3-way classification, i.e., positive, negative and neutral.Conflict data removed from SemEval-2014 datasets. The results with "♭" from [46], and those with "‡" are copied from the AEN-BERT paper [34]. "-" means not reported. For our method or re-implementations from others' code, we run the program for 10 times with random initialization, and show "mean±std" as its performance. Best and second best scores in each column are shown in bold and underlined fonts respectively.

	Text Representation	Method	Laptop-3way		Restaurant-3way		Twitter	
			Accuracy	Macro-F1	Accuracy	Macro-F1	Accuracy	Macro-F1
I	Hand-crafted Features	DCU	-	-	-	-	-	-
		NRC-Canada	-	-	-	-	-	-
		Feature-based SVM	-	-	-	-	63.40‡	67.30‡
II	Embedding	Rec-NN	-	-	-	-	66.30‡	65.90‡
		TD-LSTM	68.13‡	-	75.63‡	-	70.80‡	69.00‡
		ATAE-LSTM	68.70‡	-	77.20‡	-	-	-
		MemNet	70.33‡	64.09‡	78.16‡	65.83‡	68.50‡	66.91‡
		RAM	74.49‡	71.35‡	80.23‡	70.80‡	69.36‡	67.30‡
		IAN	72.10‡	-	78.60‡	-	-	-
		GCAE	-	-	-	-	-	-
III	BERT	AOA	77.71±1.09	73.53±1.25	80.78±0.38	69.64±1.04	73.22±0.88	73.22±0.88
		IAN	77.02±0.86	72.05±1.55	81.55±0.24	71.77±0.31	72.78±0.52	72.78±0.52
		MemNet	76.93±0.57	72.11±1.00	84.04±0.64	76.15±0.85	71.36±0.36	70.02±0.65
		RAM	77.30±0.90	72.90±1.33	83.11±0.66	74.35±1.18	74.08±0.75	72.91±0.64
IV		BERT-FC	76.54±1.42	72.83±0.99	81.28±0.60	69.79±1.41	74.64±0.23	73.04±1.15
		BERT-pair-QA-M	77.93±0.82	73.71±1.72	85.12±0.41	77.31±1.10	74.47±0.39	73.53±0.98
		AEN-BERT	78.35±1.24	73.68±1.19	81.46±0.29	71.73±1.12	73.19±0.93	71.69±0.91
		BERT-PT	78.07♭	**75.08♭**	84.95♭	76.96♭	-	-
V		TD-BERT	**78.87±1.13**	74.38±0.81	85.10±0.20	78.35±1.34	76.69±0.58	74.28±0.68
		TD-BERT-QA-MUL	78.04±0.62	73.69±0.72	**85.27±0.24**	79.15±0.86	77.04±0.45	**75.56±0.93**
		TD-BERT-QA-CON	78.42±0.15	74.37±1.39	84.56±0.50	**79.61±0.79**	**77.31±0.49**	74.40±0.39

TABLE 5. Performance comparison with classification accuracy and F1 value on the test set as the evaluation metrics. 4-way stands for 4-way classification, i.e., positive, negative, neutral and conflict. The results with "∗" are directly taken from SemEval-2014 [28], "♮" from GCAE [47]. "-" means not reported. For our method or re-implementations from others' code, we run the program for 10 times with random initialization, and show "mean±std" as its performance. Best and second best scores in each column are shown in bold and underlined fonts respectively.

	Text Representation	Method	Laptop-4way		Restaurant-4way	
			Accuracy	Macro-F1	Accuracy	Macro-F1
I	Hand-crafted Features	DCU	70.48*	-	80.95*	-
		NRC-Canada	70.48*	-	80.15*	-
		Feature-based SVM	-	-	-	-
II	Embedding	Rec-NN	-	-	-	-
		TD-LSTM	62.23±0.92♮	-	73.44±1.17♮	-
		ATAE-LSTM	64.38±4.52♮	-	73.74±3.01♮	-
		MemNet	-	-	-	-
		RAM	68.48±0.85♮	-	76.97±0.64♮	-
		IAN	68.49±0.57♮	-	76.34±0.27♮	-
		GCAE	69.14±0.32♮	-	77.28±0.32♮	-
III	BERT	AOA	74.08±0.75	54.37±1.32	80.78±0.38	59.92±1.71
		IAN	74.95±0.25	53.25±1.22	80.02±0.72	55.27±2.40
		MemNet	74.80±0.46	56.45±1.63	81.89±0.42	62.57±1.89
		RAM	74.71±0.99	56.22±1.93	82.10±1.20	61.70±1.76
IV		BERT-FC	74.93±0.88	56.95±4.21	80.87±0.20	58.93±2.77
		BERT-pair-QA-M	76.22±0.50	59.48±3.50	83.85±0.17	**65.12±1.47**
		AEN-BERT	-	-	-	-
		BERT-PT	-	-	-	-
V		TD-BERT	**76.62±0.90**	54.37±0.82	**84.37±0.28**	58.29±1.41
		TD-BERT-QA-MUL	76.53±0.68	**61.49±1.38**	84.33±0.94	64.15±1.31
		TD-BERT-QA-CON	75.83±0.70	58.17±1.79	84.15±0.48	58.98±1.99

our model both work on aspect-level sentiment classification tasks and belong to SemEval-2014 task 4. Unfortunately, [36] focuses on the solution of subtask 4 (aspect category polarity), but our task definition fits subtask 2 (aspect term polarity), so BERT-pair-QA-M is reimplemented for a fair comparison.

AEN-BERT [34] is an attention encoder network that eschews recurrence and employs attention based encoders for the modeling between context and target. From the author's published source code[4], the performance metric is calculated on the test set after every five steps, and the best performance is reported with parameter tuning on such a tight grid. To offset its unfair advantage, We modify the testing mechanism of the author's code, change the test interval to be consistent with ours (1 epoch), retrain on three data sets and report new results. The new metrics are significantly lower than in the original publication [34], especially in the Restaurant

[4]https://github.com/songyouwei/ABSA-PyTorch

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2946594, IEEE Access

collection, but we believe the results are more comparable to other work.

BERT-PT [46] assumes that aspect sentiment classification can be interpreted as a special MRC problem [32, 33], where all questions are about the polarity of a given aspect.

Experimental results are given in Table 4 and Table 5, showing overall classification accuracy and macro-F1 values in 3-way and 4-way classification, respectively. Models in part I are based on traditional feature engineering methods. Part II contains deep learning methods based on word embedding, including pre-trained word vectors (Word2Vec [21], GloVe [26]) and customized embedding vectors of its own (Rec-NN [6]). The rest (including part III, IV and V) all use BERT as the source representation, but with different network infrastructure. Networks in III are originally designed for traditional word embeddings. IV includes earlier work based on BERT. Part V contains our TD-BERT model and its variants.

We find that DCU and NRC-Canada, although not utilizing any advanced deep learning technology or embedding representation, result in strong performance. They even outperform the embedding-based methods, which demonstrates the importance of accurate feature representation for aspect-level sentiment classification. Models using BERT as input display significant improvements in classification accuracy over embedding models, indicating that BERT is indeed more capable of representing semantic and syntactic features. We also notice that the BERT-FC model, without any target information, achieves comparable performance to the models in Part III which are well-designed in the aspect-level sentiment classification task. Our hypothesis is that the previously carefully-tuned models are to strengthen the interaction between the target words and their context, in order to make up for the defect of context-free nature for the pre-trained word vectors. On the other hand, the BERT model fully considers context information of the sentence where the target word is already included in the training process. When the network based on word embeddings is combined with the BERT representation as a task-specific model, the features that can be learned are either redundant or even erroneous to a large extent. It leads to the result that their classification accuracy shows almost no improvement over the the BERT-FC model, with macro-F1 lower than BERT-FC in some cases.

Our three models achieve new state-of-the-art performance on three datasets, especially for Twitter, in which our model has a 2-3% margin over the best previous result. It shows that BERT's multi-layer bidirectional Transformer successfully encodes most of the contextual information, so that we can achieve good results by relying solely on the target. After the position output information of the target is integrated into the BERT-pair-QA-M model, the classification accuracy of TD-BERT-QA-MUL and TD-BERT-QA-CON is also improved, slightly over TD-BERT on Twitter and Restaurant in its 3-way classification task. The information fusion is applied with either element-wise multiplication or concatenation, but the performance comparison between them is almost equivalent. Although we have assumed that the introduction of auxiliary sentence will further improve the performance of TD-BERT, the difference is small and unstable, overall statistically insignificant. It might be caused by the limited domain knowledge from the small ABSA dataset, or the auxiliary sentence does not bring in any additional knowledge. It is an interesting topic for the ABSA research community how to best utilize the limited target information in a BERT-based setting, and it is what we will investigate next.

D. DISCUSSION

The experiment shows that TD-BERT significantly outperforms the BERT-pair-QA-M model on the Twitter collection, while the pattern is not observed in the other two datasets. Our assumption is that BERT-pair-QA-M model is susceptible to interference from unrelated information and works poorly for recognizing the neutral polarity. Negative, neutral, positive samples account for 25%, 50%, 25%, respectively, in the Twitter datasets, and the percentage of neutral samples is much smaller in the other two datasets, in which BERT-pair-QA-M works comparably well.

In order to verify the assumption, we take subsets of the original data, consisting of the neutral samples in the test set of the Laptop, Restaurant, and Twitter collections. By training on the original training set and evaluating on the new test set, we can see how well these models work on the neutral cases. Experiment results are shown in Table 6.

We can see that the TD-BERT method has clear advantages over BERT-pair-QA-M for neutral data. In terms of classification accuracy, TD-BERT is 3-11% higher, especially on the laptop dataset. Naturally, it also leads to the conclusion that the classification accuracy of TD-BERT on positive and negative cases may be slightly lower, which is worth further analysis. How to combine their strengths so that we can achieve stable improvements on all classes? That is what we need to find out.

TABLE 6. The accuracy of different models on neutral samples only. Higher score in each row is marked with bold font.

Method	TD-BERT	BERT-pair-QA-M
Laptop	**69.42±3.84**	58.18±3.75
Restaurant	**57.65±2.52**	54.59±3.03
Twitter	**83.09±1.19**	80.27±3.34

To further examine our model in complex cases, we test its expressiveness in a multi-target scenario with inconsistent sentiment polarities. We construct a hard-data-only dataset, which is described in Table 2. The test results are shown in Table 7. In terms of classification accuracy, the TD-BERT model is 6-10% higher than BERT-pair-QA-M, showing its advantage in handling complex sentiment labels with multiple targets.

We believe that a multi-target task is more capable of reflecting the difficulties faced by fine-grained sentiment analysis, and it deserves more in-depth research. We also

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2946594, IEEE Access

agree with [8] that correlation and influence exist among multiple targets in a single text piece, but similar work has been rarely observed in this area. The reason might be that previous models are already complex enough for single-target tasks, and extending it to an inter-correlated multi-target scenario will bring too much challenge. Fortunately, our TD-BERT model is originated from a straightforward idea, and is easy to extend to multi-target cases. We plan to consider the interaction between different targets in an extended TD-BERT model next.

TABLE 7. The accuracy of different models on hard data only. Higher score in each row is marked with bold font.

Method	TD-BERT	BERT-pair-QA-M
Laptop	**53.21±1.14**	47.17±1.60
Restaurant	**48.61±4.10**	39.93±2.44

V. CONCLUSION

BERT has displayed its great advantage of text representation in many NLP tasks, including sentence-level sentiment classification. However, its application to sentiment analysis at the aspect level is rare. In this paper, We explore its representative power in target-dependent sentiment classification, which is a subtask of ABSA. Well-designed feature engineering with a good classifier still outperforms deep learning models with a word embedding representation, but BERT has raised the baseline of the game to a totally different level. Those complex models customized for embeddings do not work well with BERT, sometimes even below the vanilla implementation of BERT representation (BERT-FC). Earlier work with BERT has shown noticeable improvements over its strong baseline, some with simple ideas (BERT-pair-QA-M), others with a more complex structure (AEN-BERT) or even with additional datasets and tasks (BERT-PT). Our implementation is mainly based on a small revision to focus on the target terms instead of the whole sentence. Together with some variants that form an auxiliary sentence with the target, they establish new state-of-the-art on SemEval-2014 and a Twitter dataset.

It also comes to our attention that the improvements over BERT baseline, although statistically significant, do not resemble the 5-10% or higher boost we used to see in embedding-based models. Does it mean that we still have not identified the appropriate network structure to exhaust the potential of BERT representation, or it has set a baseline so high that there is not much room for improvement at all? This is an interesting research question, which we plan to continue working on. At this time, our preference goes to the latter, as existing work with BERT has exhibited similar patterns.

Another observation is that the average classification accuracy has been pushed to high 70s or mid 80s in percentage, but there are still certain classes of data, for which the current model cannot provide a satisfactory solution. The classification accuracy of neutral cases is much lower that those with a clear polarity, and those with mixed sentiment polarities towards different aspects (hard data) or the same target (conflict) are even harder to process. Accurate identification of such cases requires more training data, together with in-depth analysis to extract useful patterns. This is a more challenging task that we plan to tackle next.

REFERENCES

[1] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473, 2014.

[2] Yoshua Bengio, Réjean Ducharme, Pascal Vincent, and Christian Jauvin. A neural probabilistic language model. Journal of machine learning research, 3(Feb): 1137–1155, 2003.

[3] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. Journal of machine Learning research, 3(Jan):993–1022, 2003.

[4] Peng Chen, Zhongqian Sun, Lidong Bing, and Wei Yang. Recurrent attention network on memory for aspect sentiment analysis. In Proceedings of the 2017 conference on empirical methods in natural language processing, pages 452–461, 2017.

[5] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805, 2018.

[6] Li Dong, Furu Wei, Chuanqi Tan, Duyu Tang, Ming Zhou, and Ke Xu. Adaptive recursive neural network for target-dependent twitter sentiment classification. In Proceedings of the 52nd annual meeting of the association for computational linguistics (volume 2: Short papers), volume 2, pages 49–54, 2014.

[7] Alex Graves, Greg Wayne, and Ivo Danihelka. Neural turing machines. arXiv preprint arXiv:1410.5401, 2014.

[8] Devamanyu Hazarika, Soujanya Poria, Prateek Vij, Gangeshwar Krishnamurthy, Erik Cambria, and Roger Zimmermann. Modeling inter-aspect dependencies for aspect-based sentiment analysis. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers), pages 266–270, 2018.

[9] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. Neural computation, 9(8):1735–1780, 1997.

[10] Thomas Hofmann. Probabilistic latent semantic analysis. In Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence, pages 289–296. Morgan Kaufmann Publishers Inc., 1999.

[11] Jeremy Howard and Sebastian Ruder. Universal language model fine-tuning for text classification. arXiv preprint arXiv:1801.06146, 2018.

[12] Minqing Hu and Bing Liu. Mining and summarizing customer reviews. In Proceedings of the tenth ACM

SIGKDD international conference on Knowledge discovery and data mining, pages 168–177. ACM, 2004.

[13] Binxuan Huang and Kathleen Carley. Parameterized convolutional neural networks for aspect level sentiment classification. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pages 1091–1096, 2018.

[14] Binxuan Huang, Yanglan Ou, and Kathleen M Carley. Aspect level sentiment classification with attention-over-attention neural networks. In International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation, pages 197–206. Springer, 2018.

[15] Long Jiang, Mo Yu, Ming Zhou, Xiaohua Liu, and Tiejun Zhao. Target-dependent twitter sentiment classification. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1, pages 151–160. Association for Computational Linguistics, 2011.

[16] Nal Kalchbrenner, Edward Grefenstette, and Phil Blunsom. A convolutional neural network for modelling sentences. arXiv preprint arXiv:1404.2188, 2014.

[17] Yoon Kim. Convolutional neural networks for sentence classification. arXiv preprint arXiv:1408.5882, 2014.

[18] Svetlana Kiritchenko, Xiaodan Zhu, Colin Cherry, and Saif Mohammad. Nrc-canada-2014: Detecting aspects and sentiment in customer reviews. In Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), pages 437–442, 2014.

[19] Bing Liu. Sentiment analysis and opinion mining. Synthesis lectures on human language technologies, 5 (1):1–167, 2012.

[20] Dehong Ma, Sujian Li, Xiaodong Zhang, and Houfeng Wang. Interactive attention networks for aspect-level sentiment classification. arXiv preprint arXiv:1709.00893, 2017.

[21] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781, 2013.

[22] Alena Neviarouskaya, Helmut Prendinger, and Mitsuru Ishizuka. Sentiful: Generating a reliable lexicon for sentiment analysis. In 2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops, pages 1–6. IEEE, 2009.

[23] Thien Hai Nguyen and Kiyoaki Shirai. Phrasernn: Phrase recursive neural network for aspect-based sentiment analysis. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pages 2509–2514, 2015.

[24] Bo Pang, Lillian Lee, et al. Opinion mining and sentiment analysis. Foundations and Trends® in Information Retrieval, 2(1–2):1–135, 2008.

[25] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. 2017.

[26] Jeffrey Pennington, Richard Socher, and Christopher Manning. Glove: Global vectors for word representation. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), pages 1532–1543, 2014.

[27] Matthew E Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. Deep contextualized word representations. arXiv preprint arXiv:1802.05365, 2018.

[28] Maria Pontiki, Dimitris Galanis, John Pavlopoulos, Harris Papageorgiou, Ion Androutsopoulos, and Suresh Manandhar. Semeval-2014 task 4: Aspect based sentiment analysis. In Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), pages 27–35, 2014.

[29] Ye Qi, Devendra Singh Sachan, Matthieu Felix, Sarguna Janani Padmanabhan, and Graham Neubig. When and why are pre-trained word embeddings useful for neural machine translation? arXiv preprint arXiv:1804.06323, 2018.

[30] Guang Qiu, Liu Bing, Jiajun Bu, and Chun Chen. Expanding domain sentiment lexicon through double propagation. In International Jont Conference on Artifical Intelligence, 2009.

[31] Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. Improving language understanding by generative pre-training. URL https://s3-us-west-2. amazonaws. com/openai-assets/research-covers/languageunsupervised/language understanding paper. pdf, 2018.

[32] Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. Squad: 100,000+ questions for machine comprehension of text. arXiv preprint arXiv:1606.05250, 2016.

[33] Pranav Rajpurkar, Robin Jia, and Percy Liang. Know what you don't know: Unanswerable questions for squad. arXiv preprint arXiv:1806.03822, 2018.

[34] Youwei Song, Jiahai Wang, Tao Jiang, Zhiyue Liu, and Yanghui Rao. Attentional encoder network for targeted sentiment classification. arXiv preprint arXiv:1902.09314, 2019.

[35] Sainbayar Sukhbaatar, Jason Weston, Rob Fergus, et al. End-to-end memory networks. In Advances in neural information processing systems, pages 2440–2448, 2015.

[36] Chi Sun, Luyao Huang, and Xipeng Qiu. Utilizing bert for aspect-based sentiment analysis via constructing auxiliary sentence. arXiv preprint arXiv:1903.09588, 2019.

[37] Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly Voll, and Manfred Stede. Lexicon-based methods for sentiment analysis. Computational Linguistics, 37 (2):267–307, 2011.

[38] Duyu Tang, Bing Qin, Xiaocheng Feng, and Ting Liu. Effective lstms for target-dependent sentiment classifi-

cation. arXiv preprint arXiv:1512.01100, 2015.

[39] Duyu Tang, Bing Qin, and Ting Liu. Aspect level sentiment classification with deep memory network. arXiv preprint arXiv:1605.08900, 2016.

[40] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In Advances in neural information processing systems, pages 5998–6008, 2017.

[41] Duy-Tin Vo and Yue Zhang. Target-dependent twitter sentiment classification with rich automatic features. In Twenty-Fourth International Joint Conference on Artificial Intelligence, 2015.

[42] Joachim Wagner, Piyush Arora, Santiago Cortes, Utsab Barman, Dasha Bogdanova, Jennifer Foster, and Lamia Tounsi. Dcu: Aspect-based polarity classification for semeval task 4. In Proceedings of the 8th international workshop on semantic evaluation (SemEval 2014), pages 223–229, 2014.

[43] Bo Wang, Maria Liakata, Arkaitz Zubiaga, and Rob Procter. Tdparse: Multi-target-specific sentiment recognition on twitter. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers, pages 483–493, 2017.

[44] Yequan Wang, Minlie Huang, Li Zhao, et al. Attention-based lstm for aspect-level sentiment classification. In Proceedings of the 2016 conference on empirical methods in natural language processing, pages 606–615, 2016.

[45] Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V Le, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikun, Yuan Cao, Qin Gao, Klaus Macherey, et al. Google's neural machine translation system: Bridging the gap between human and machine translation. arXiv preprint arXiv:1609.08144, 2016.

[46] Hu Xu, Bing Liu, Lei Shu, and Philip S Yu. Bert post-training for review reading comprehension and aspect-based sentiment analysis. arXiv preprint arXiv:1904.02232, 2019.

[47] Wei Xue and Tao Li. Aspect based sentiment analysis with gated convolutional networks. arXiv preprint arXiv:1805.07043, 2018.

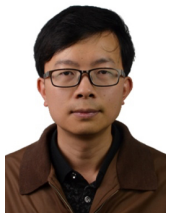
[48] Lei Zhang, Shuai Wang, and Bing Liu. Deep learning for sentiment analysis: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8(4):e1253, 2018.

[49] Meishan Zhang, Yue Zhang, and Duy-Tin Vo. Gated neural networks for targeted sentiment analysis. In Thirtieth AAAI Conference on Artificial Intelligence, 2016.

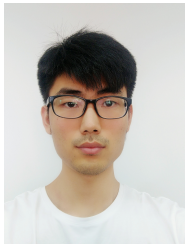
[50] Chunting Zhou, Chonglin Sun, Zhiyuan Liu, and Francis Lau. A c-lstm neural network for text classification. arXiv preprint arXiv:1511.08630, 2015.

[51] Peisong Zhu and Tieyun Qian. Enhanced aspect level sentiment classification with auxiliary memory. In Proceedings of the 27th International Conference on Computational Linguistics, pages 1077–1087, 2018.

ZHENGJIE GAO received the B.E. degree in computer science and technology from Chengdu University of Information Technology, Chengdu, China, in 2017, where he is currently pursuing the M.E. degree in computer application technology. His research interests include sentiment analysis, text classification, and deep learning.



AO FENG received the B.E. and M.E. degrees in Automation from Tsinghua University, China in 1999 and 2001, respectively. He received his M.S. and Ph.D. degrees in Computer Science from University of Massachusetts Amherst, U.S. in 2008. He has worked at Amazon.com and Lenovo research, and is currently an associate professor at the Department of Computer Science, Chengdu University of Information Technology. His research interests include information retrieval, data mining, natural language processing, and machine learning.



XINYU SONG received the B.E. degree from Heilongjiang Bayi Agricultural University, Daqing, China, in 2017, where he is currently pursuing the M.E. degree in computer technology in Chengdu University of Information Technology. His research interests include sentiment analysis, information extraction and deep learning.

XI WU received his B.S. degree in Communication Engineering from Sichuan University in 2003, M.S. degree in Communication Engineering from University of Electronic Science and Technology of China in 2006, and Ph.D. degree in Communication Engineering from Sichuan University in 2009. He is currently a professor in the Department of Computer Science, Chengdu University of Information Technology. His current research interests include image processing and computer vision.