BERT-Based Multi-Task Learning for Aspect-Based Opinion Mining

Manil Patel(✉) and C. I. Ezeife(✉)

School of Computer Science, University of Windsor, 401 Sunset Avenue, Windsor, ON N9B3P4, Canada
{patel3h,cezeife}@uwindsor.ca

Abstract. Aspect-Based Opinion Mining (ABOM) mainly focuses on mining the aspect terms (product's features) and related opinion polarities (e.g., Positive, Negative, and Neutral) from user's reviews. The most prominent neural network-based methods to perform ABOM tasks include BERT-based approaches, such as BERT-PT and BAT. These approaches build separate models to complete each ABOM subtasks, such as aspect term extraction (e.g., pizza, staff member) and aspect sentiment classification. Both approaches use different training algorithms, such as Post-Training and Adversarial Training. Also, the BERT-LSTM/Attention approach uses different pooling strategies on the intermediate layers of the BERT model to achieve better results. Moreover, they do not consider the subtasks of aspect categories (e.g., a category of aspect pizza in a review is food) and related opinion polarity. This paper proposes a new system for ABOM, called BERT-MTL, which uses Multi-Task Learning (MTL) approach and differentiates from these previous approaches by solving two tasks such as aspect terms and categories extraction simultaneously by taking advantage of similarities between tasks and enhancing the model's accuracy as well as reduce the training time. Our proposed system also builds models to identify user's opinions for aspect terms and aspect categories by applying different pooling strategies on the last layer of the BERT model. To evaluate our model's performance, we have used the SemEval-14 task 4 restaurant dataset. Our model outperforms previous models in several ABOM tasks, and the experimental results support its validity.

Keywords: Aspect-based opinion mining · BERT · Multi-task learning · Sentiment analysis · Pooling strategies

1 Introduction

Opinion mining aims to extract people's opinions or sentiments (e.g., positive or negative) and subjectivity (subjective statements are those statements which

This research was supported by the Natural Science and Engineering Research Council (NSERC) of Canada under an Operating grant (OGP-0194134) and a University of Windsor grant.

C. Strauss et al. (Eds.): DEXA 2021, LNCS 12923, pp. 192–204, 2021.
https://doi.org/10.1007/978-3-030-86472-9_18

contain opinion terms) from the texts [4]. There are large numbers of user's opinions available about a particular place (e.g., Hotel, Restaurant) and electronic products (e.g., Laptop, Phone) on different applications (e.g., TripAdvisor, Amazon). It is exceedingly difficult for a user to review/read all these available opinions and decide whether to visit a place or not, buy a product or not.

The Aspect-Based Opinion Mining (ABOM) focuses on the aspect term and term related opinion polarity (positive, negative, and neutral) [14]. In other words, instead of classifying the general opinion of the text as positive or negative, aspect-based analysis allows one to associate strong opinions with specific features of the product or service [4]. For example, "The pizza was delicious, but the staff members were horrible to us." The ABOM system detects opinion polarity for each aspect term (pizza = positive, staff members = negative). The ABOM task consists of four subtasks such as Aspect Term Extraction (ATE), Aspect Term Related Polarity (Fine-grained ABOM), Aspect Category Detection (ACD), and Aspect Category Related Polarity (Coarse-grained ABOM).

From the given sentence as an example, "The pizza here is rather good, but only if you like to wait for it." we can get aspect term such as pizza from the example, but the sentence does not have any aspect terms related to service as it shows the only context related to service. However, the aspect category model can detect food and service both as it depends on context rather than the aspect term, due to which finding of aspect terms and categories is essential while performing ABOM.

BERT [3] stands for Bidirectional Encoder Representations from Transformers. We are using the pre-trained BERT-BASE model, which contains 12 encoder layers, and each encoder layer contains Attention, Layer Norm, and Feed Forward Neural Network (FFNN). Unlike the other neural network-based approaches, BERT is pre-trained, which can be fine-tuned with just one additional output layer. The basic idea behind pre-training is that it trained the model by different datasets and used their weights as an initial weight in the model [16]. Therefore, pre-training gives the network a head start. Also, BERT uses the tokenize function to generate the tokens and each token-related ids from the input sentences. BERT tokenize function adds two unique tokens in each sentence, such as [CLS] and [SEP]. The token [CLS] is used to indicate the starting of sequences and classification, while [SEP] is used to separate the sequence from the subsequent. To generate Tokens, Input_Id, Attention_Mask, token_type_ids from a sentence, the BERT Tokenizer function converts each sentence into a list of tokens. Input_Id are token indices (numerical representations of tokens), and Attention_Mask is used to identify the tokens and padding, where tokens are represented as one and padding denotes Zero. token_type_ids are required two different sequences to be joined in a single "input_ids" entry, where the "context" corresponding to the question in the first sequence is represented by a 0, whereas the "question" in the second sequence is represented by a 1 [3].

1.1 Problem Definition

The user's reviews contain multiple aspect terms and categories as well as their related opinion polarity, which is essential to identify the user's actual concerns on specific features. Previous research (BERT-PT [17], BAT [8], BERT-LSTM/Attention [15]) require individual models to perform each subtask of ABOM (e.g., ATE and Fine-grained ABOM). The problem identified in this research is to extract the aspect terms and aspect categories for each user's reviews with minimum computation and at the same time.

1.2 Contributions

1. Our approach (BERT-MTL) can extract aspect categories and aspect terms simultaneously using Multi-Task Learning, which will identify the commonalities and differences between the tasks to enhance model's performance.
2. Also, our approach trained on a half number of the epochs as compared to previous BERT based approaches (BERT-PT [17], BAT [8], BERT-LSTM/Attention [15]) due to which we can say that our proposed approach required less amount of time to train on data.
3. In our experiments, we find that the pooling strategies work better on the final layer of BERT than the intermediate layers of the BERT model, and the results of the model validate the claim.
4. The proposed BERT-based approach (BERT-MTL) achieves better results on every subtask of ABOM compared to the previous State-of-the-art model.

The rest of the paper is summarized as follows; Sect. 2 presents Literature Review. Section 3 is the Proposed Approach called BERT-MTL. In Sect. 4, the Steps of the Proposed Algorithm are discussed, and in Sect. 5, Experimental Evaluation is discussed. In the end, in Sect. 6, the Conclusions and Future Work are presented.

2 Literature Review

Some of the past state-of-the-art approaches in the ABOM are given below.

MGAN: The Multi Granularity Alignment Network (MGAN) [10] can perform the Coarse-grained ABOM and Fine-grained ABOM. To achieve both the task, the author's designed the Coarse2Fine Attention module, which can transfer the aspect knowledge to the coarse-grained and fine-grained networks.

IMN: The authors [5] proposed Interactive Multi-Task Learning Network (IMN), which can jointly perform aspect and opinion term extraction as well as Fine-grained ABOM using Convolutional Neural Network (CNN) and Attention model. It can also learn from multiple training data to exploit the correlation between Fine-Grained ABOM and Document-level Opinion Mining.

BERT-Post Training: The authors [17] developed three different BERT-based models to perform the ABOM task. In the first model (ATE), they perform Named Entity Recognition (NER) to extract the aspect terms from the user reviews using BERT. The second model (ATSC) performs a classification task to predict opinion polarity for the aspect terms. The last model (RRC) is unique, and it is based on question answering where the user asked a question, and the BERT [3] model will predict answers from the user reviews. Also, they proposed Post Training algorithm to train all three BERT-based approaches.

AAL: The Aspect Aware Learning (AAL) model [21] uses three components to perform the Coarse-grained ABOM task. The first component (Network input) converts the words in a sentence to their respective vector form using Word2vec, and Bi-LSTM [6]. The second component (AAL) captures the correlation between aspect term and category. The third component (sentiment classification) performs the Coarse-grained ABOM using the attention layer.

BERT-LSTM/Attention: The authors [15] used intermediate layers of BERT and design two knowledge pooling strategies such as LSTM and Attention-based to perform ABOM task on different datasets. The authors [15] applied the post-training concept on the proposed approached and during evaluation, the post-training based BERT-LSTM/Attention achieved higher results.

BAT: BERT Adversarial Training (BAT) approach [8], first create the adversarial example by applying small perturbations to the original inputs. Although these examples are not actual sentences, they have been shown to serve as a regularization mechanism that can enhance the robustness of neural networks. The adversarial examples were then fed into the BERT encoder model for training, and the model was able to achieve high accuracy with the aid of the adversarial training model.

3 Proposed Approach

The previous BERT-based approaches require different models to perform all the four subtasks of ABOM from user reviews. To resolve that problem, we suggested a novel BERT-based model (BERT-MTL) for ABOM. In which, we have build three BERT-based models. The First model performs BERT-based Multi-Task Learning (MTL) approach to extract the aspect terms and aspect categories from the user reviews. The second model performs the Fine-grained ABOM, and the third model performs the Coarse-grained ABOM using different pooling straties on last layer of the BERT model.

3.1 Multi-Task Learning Model

To extract all the aspect terms and aspect categories present in the user reviews, we will use Multi-Task Learning (MTL) approach using BERT. In MTL, various tasks are learned together in a single network, each task having its own output [1]. MTL captures the similarities between tasks and improves model generalization capacity in certain situations by learning semantically similar tasks in parallel using a shared representation [5]. The BERT-MTL network has a common input, and layers of the BERT model are shared between the two tasks, such as Aspect Term Extraction (ATE) and Aspect Category Detection (ACD).

We performed two different methods in the BERT-MTL model, such as Sequence Labeling to extract the aspect terms and Multi-Label Classification to identify aspect categories. We have build a model which can take user reviews as an input, and it can predict aspect term and aspect category at the same time from those reviews by sharing the model knowledge, which means if the model learns about the aspect term, then it will help the aspect category task to learn, and vice-versa [1]. For example, If the user's review contains the name of a dish (e.g., pizza), it is easy to infer from the context that it is about the aspect category Food.

According to the previous research, the use of MTL reduces the chances of overfitting the model, which is a bigger problem in most neural network approaches[13]. Our proposed approach showed in Algorithm 1; by giving user review as an input to the model, it can predict the aspect term and aspect category simultaneously by applying MTL. In the BERT-MTL approach, we calculate the final loss function as sum of both the task-related loss function such as $Loss = Loss_{ACD} + Loss_{ATE}$. where $Loss_{ACD}$ calculated by BCEWithLogitsLoss() function and $Loss_{ATE}$ calculated by CrossEntropyLoss() [20] function.

Algorithm 1: BERT-MTL for ATE and ACD

Input: Training Sentences (s) from Dataset
Output: Aspect Terms and Categories
Initialize: Initialize the hyper-parameter learning rate, dropout probability
Loop until the terminal condition is met. Maximum Training Epochs:

- $Sentences_{batch} \leftarrow sample(Sentences; b)$; `// sample a minibatch of size b`
- $Input\_Id, Attention\_Mask, token\_type\_ids = BERT\ Tokenizer(Sentences)$; `//`
- $x_i = BERT\ Embedding\ Function(Input\_Id, Attention\_Mask, token\_type\_ids)$; `//`
- $Final, \_ = BERT\ Encoder\ Function(x_i)$; `// Output from last layer of BERT`
- $h_t = Pooling\ Strategy(Final)$; `// LSTM and GRU NN`
- $prediction1(ATE) = Feed - Forward\ NN(h_t)$; `// Classification layer 1`
- $prediction2(ACD) = Feed - Forward\ NN(h_t)$; `// Classification layer 2`
- $Loss1 = CrossEntropyLoss(prediction1, target1)$; `// Loss for ATE`
- $Loss2 = BCEWithLogitsLoss(prediction2, target2)$; `// Loss for ACD`
- $Loss = Loss1 + Loss2$; `// sum the Loss`
- $Back - propagation\ algorithm\ is\ used\ to\ changed\ the\ weights\ of\ the\ approach$

end

Sequence Labeling (ATE (Task 1)): In Sequence Labeling, each word in the sentence has a label in the BIO format. Where B stands for the Beginning of aspect terms, I stands for Inside (continue) of aspect terms, and O stands for Outside of aspect terms. Some of the aspect terms are phrase level (two or more words, e.g., Cheese Garlic Bread), due to which we need the Inside label to identify all the words that can be considered as aspect terms. For the given sentence "Garlic Bread is good.", the output of sequence labeling task for each word of input will be Garlic = B, Bread = I, is = O, and good = O.

Multi-Label Classification (ACD (Task 2)): In multi-label classification, each sentence has its list of labels based on how many categories are displayed in one sentence. In our case, we consider that the sentence can be classified into five aspect categories as Service, Food, Price, Ambiance, and Miscellaneous. For the given sentence "pizza is good, but staff members were horrible to us.", the output of the multi-Label classification task for input will be Food and Service, as it indicates context related to these categories.

3.2 Aspect Term and Category Related Opinion Polarity

Fine-Grained ABOM: After extracting the aspect term and aspect category from the user's reviews, we aim to determine the opinion polarity (Positive, Negative, and Neutral) related to each of the aspect terms and categories. We will use the sentence pair classification approach using BERT [3] model to perform that task. In this task, we will give user review and aspect term as an input to the model (e.g., (user review, aspect term) a pair given as input) due to which model can predict the opinion related to that aspect from a review.

In sentence pair classification, the input for the BERT model will be sentenced aspect pair as represented below:

$$\begin{aligned} Input_2 &= ([CLS], w_1, w_2, ..., w_n, [SEP], a_1, a_2, ..., a_n, [SEP]);\\ Final, \_ &= BERT(Input_2);\\ h_t &= Pooling\ Strategy(Final);\\ Y &= softmax(W_3 \cdot h_t + b_3); \end{aligned}$$

where $w_1, w_2, w_3, ..., w_n$ are words in a sentence which contains the aspect terms while a_1, a_2, ..., a_n are the aspect term present in a sentence. When a sentence contains multiple aspect terms at that time sentence is repeated with each unique aspect term present in a sentence as a pair of input to the BERT model. The weight $W_3 \in \mathbb{R}^{3*d_I}$ and bias $b_3 \in \mathbb{R}^3$ (3 is a total number of classes (Positive, Negative, and Neutral) and d_I is a hidden dimension of Pooling Strategy).

Coarse-Grained ABOM: To perform the Coarse-grained ABOM, we perform the same task as we performed for Fine-grained ABOM. We are passing sentence and aspect category to detect opinion polarity related to that aspect category in input.

$$\begin{aligned} Input_3 &= ([CLS], w_1, w_2, ..., w_n, [SEP], c_1, c_2, ..., c_n, [SEP])\\ Final, \_ &= BERT(Input_3);\\ h_t &= Pooling\ Strategy(Final);\\ Y &= softmax(W_4 \cdot h_t + b_4); \end{aligned}$$

where w_1, w_2, w_3, ..., w_n are words in a sentence which contains the aspect categories or related words while c_1, c_2, ..., c_n are the aspect category present in a sentence. When a sentence contains multiple aspect categories, the sentence is repeated with each unique aspect category present in a sentence as a pair of inputs to the BERT model. The computation will be same as Fine-grained ABOM approach.

Pooling Strategy: In this paper, we applied three different network-based pooling strategies, such as Dot Product Attention [11], LSTM [6], and GRU [2], on the last layer of the BERT model. As LSTM and GRU networks are suitable for processing sequence information, we used both the network to connect the last layer of the BERT model [15]. The last cell of the LSTM and GRU will be the output of the network, which will be given to the FFNN layer to predict the output.

LSTM Pooling Strategy: The LSTM pooling strategy applied on the last layer of BERT are discussed below:

$$Final, \_ = BERT(Input_i)$$
$$h_t = \overrightarrow{LSTM}(Final)$$

where $Input_i$ in BERT denote either input is $Input_1$ (ATE and ACD), $Input_2$ (Fine-grained ABOM) or $Input_3$ (Coarse-grained ABOM). Also, h_t represents the LSTM network output.

GRU Pooling Strategy: The GRU pooling strategy applied on the last layer of BERT are discussed below:

$$Final, \_ = BERT(Input_i)$$
$$h_t = \overrightarrow{GRU}(Final)$$

where $Input_i$ in BERT denote either input is $Input_1$ (ATE and ACD), $Input_2$ (Fine-grained ABOM) or $Input_3$ (Coarse-grained ABOM). Also, h_t represents the GRU network output.

Attention Pooling Strategy: We have also used the dot product attention on the last layer of the BERT model as a pooling strategy. The computational formula for the attention layer will be given below:

$$Final, \_ = BERT(Input_i)$$
$$h_t = W_h softmax(a \cdot Final^T) Final$$

where $Input_i$ in BERT denote either input is $Input_2$ (Fine-grained ABOM) or $Input_3$ (Coarse-grained ABOM). w_h and a are the learnable parameters of the Attention network, and softmax is the activation function. The output (h_t) generated by all the three pooling strategies will be given to the FFNN layer as input to generate the final prediction probability.

4 Steps of BERT-MTL

4.1 Steps of BERT-MTL for ATE and ACD Algorithm

1. Dataset contains sentences or reviews, and each sentence has a target. For example, "The pizza is good." and the target value for a sentence is "pizza" and "food" to determine the aspect term and category, respectively.
2. First, the sentences are given to the BERT Tokenize Function to generate the tokens, Input_Id, Attention_Mask, Token_Type_Ids. For example, from above sentence, (Tokens = ['[CLS]', 'the', 'pizza', 'is', 'good', '.', '[SEP]']), (Input_Id = [101, 1996, 10,733, 2003, 2204, 1012, 102]), (Attention_Mask = [1, 1, 1, 1, 1, 1, 1]), (token_type_ids = [0, 0, 0, 0, 0, 0, 0]).
3. The above generated Input_Id, Attention_Mask, and token_type_ids will be given in BERT Embedding Function to convert each token into respective vectors called as embedding vector. After that the embedding vector related to each token fed to the BERT Encoder Function which will generate a 768 dimension vector for each token of the sentence.
4. Each token related to a 768 dimension vector will be given to the pooling strategy (discussed in above Sect. 3.2). The pooling strategy is another neural network attached to the BERT Encoder function to identify the context and generate the final result.
5. The output of the pooling strategy given to the two different FFNN to perform the different tasks, such as sequence labeling to extract aspect terms and Multi-label classification to detect aspect categories, where it converts the vectors into each class probabilities. For example, class probability generated by FFNN for word pizza in task ATE will be [0.6, 0.1, 0.2] and for sentence in task ACD will be [0.15, 0.6, 0.2, 0.05, 0.17].
6. The class probabilities generated by each FFNN and their respective actual target from the dataset will be given to the loss function. For example, actual target in task ATE for word pizza will be [1,0,0] (1 at first position indicate Beginning of the aspect term) and for ACD [0,1,0,0,0] (1 at second position indicate Food as aspect category). The loss will be calculated for both the task as displayed below:

$$Loss_{ATE} = -\sum t_i \cdot log(p_i)$$

$$Loss_{ACD} = -w_n[t_i \cdot log(\sigma(p_i)) + (1 - t_i) \cdot log(\sigma(1 - p_i))]$$

   where w_n denote the weights by default its set to none, t_i denotes target value and p_i denotes the predicted value.
7. After the calculation of loss for two different tasks such as ATE and ACD, both loss will be sum to generate the final loss function. $Loss = Loss_{ATE} + Loss_{ACD}$. After that, to reduce the final loss from both the tasks, the back-propagation algorithm was used to change the weights of the approach.

4.2 Steps of BERT-MTL for Fine and Coarse-Grained Algorithm

1. Dataset contains sentences or reviews with aspect, and each (sentence, aspect) pair have targets. For example, ("The pizza is good.", pizza/food) pair will be

input, and the target value for a pair is "positive" to determine the opinion polarity from the sentence and aspect.

2. First, the (sentence,aspect) pair is given to the BERT Tokenize Function to generate the tokens, Input_Id, Attention_Mask, and token_type_ids. And Later, it will be processed through the BERT Embedding and Encoder function and generate a 768 dimension vector for each token of the sentence.
3. Each token related a 768 dimension vector generated by BERT Encoder would be given to the pooling strategy. The pooling strategy is another neural network attached to the BERT Encoder function to identify the context and generate the final result.
4. The output of the pooling strategy given to the FFNN to perform the classification task, such as Multi-class classification to detect aspect term/category related opinion polarity, where it convert the vectors into class probabilities.
5. The class probabilities probability generated by each FFNN and their respective actual target from the dataset will be given to the BCEWithLogitsLoss function, where a loss will be calculated. Then, to reduce the loss of the task, the backpropagation algorithm was used.

5 Experimental Evaluation

In this section, we discuss the experiment and its results in detail. We tested our model in Google Colab - GPU[1]. We used the BERT-based Aspect-Based Sentiment Analysis (ABSA) code created by Avinash Sai[2] and performed changes in the code according to our approach. We implemented our code in NumPy 1.19.5, PyTorch 1.7.1, and Hugging Face transformers 4.3.2 environment.

5.1 Dataset

We conduct experiments on the SemEval-2014 task 4 [12] dataset, which contains customer reviews on restaurants. The statistics related to the dataset are represented in Table 1 for each task. In Table 1, each number in the Train and Test row represents the number of user reviews present in the dataset for every task. We have removed the sentences from the dataset, which leads to the conflict opinion polarity because the number of user reviews is small with conflict opinion polarity. The Fine-grained ABOM (training and testing combined) dataset contains 2892 positive, 1001 negative, and 829 neutral sentences, while Coarse-grained ABOM (training and testing combined) data contains 2836 positive, 1061 Negative, and 594 Neutral sentences. In the dataset, not every review contains the aspect terms due to which it may affect the statistics of the data related to Fine-grained ABOM and Coarse-grained ABOM tasks. To evaluate our model's performance, we will consider two evaluation strategies: Accuracy and F1-score. In evaluation, we are using the Macro F1-score because it is used to deal with the problem of unbalanced class, and Macro F1-score is calculated as the average F1-score of each class [8].

1 https://colab.research.google.com.
2 https://github.com/avinashsai/BERT-Aspect.

Table 1. Statistics of the dataset

Dataset	ATE and ACD	Fine-grained ABOM	Coarse-grained ABOM
Train	3,044	3,602	3518
Test	800	1,120	973

5.2 Hyper-Parameters

The selection of hyper-parameters is essential during the evaluation of model performance. We have used BERT-BASE uncased model to conduct experiments on the dataset. We have executed our models several times to figure out what number of epochs and how much dropout probability yields the highest results for our approach. Epochs indicate how many iterations are required by model to train on dataset. After the execution of the models on different parameter, we found that the dropout probability and epochs for the proposed approach should be 0.3 and 4, respectively in all three approaches. Also, we have used the Adam optimizer for better learning, and the learning rate is set to be $2e-5$ in all three approaches. During training, batch size is set to 16 and during testing it is 4 in ATE and ACD tasks and 8 in Fine and Coarse-grained ABOM tasks.

5.3 Result Analysis

In this section, we compare our results with several state-of-the-art models on the SemEval-2014 task 4 restaurant dataset. Also, not all the models include the aspect category and aspect term related opinion polarity, due to which comparison is based on every task. The result for each subtask, such as Aspect Term Extraction (ATE), Aspect Category Detection (ACD), Fine-grained ABOM, and Coarse-grained ABOM, are displayed in Tables 2, 3a, and 3b, respectively. In all the comparisons of the tables ACC represents the accuracy. The higher value of accuracy (ACC) and Macro - F1 denotes the better model. As we can see in Table 2, the BERT-MTL achieve good results on ACD and ATE tasks in terms of Macro-F1. Also, BERT-MTL with GRU pooling strategy achieves higher results on ATE task compared to other methods. Our approach outperforms most of the previous BERT-based models such as BERT-PT, BAT, DomBERT, and BERT-LSTM/Attention in Fine-grained ABOM with three different pooling strategies. Also, All these BERT-based approaches trained their models up to 10 epochs while our models train up to only 4 epochs, due to which we achieve a better result with less amount of computation. In Coarse-grained ABOM, our approach BERT-MTL achieve state-of-the-art results with all the three pooling strategies and outperforms all the previous approaches in terms of accuracy and Macro-F1.

Table 2. Result of ACD and ATE

Model	ACD (Macro F1)	ATE (Macro F1)	Epochs
MTNA [20]	88.91	84.01	-
NRC-Canada [9]	88.58	80.19	-
BERT-PT [17]	-	77.97	10
IMN [5]	-	84.01	-
DomBERT [18]	-	77.21	10
BERT-MTL	**90.18 (± 1.00)**	**84.86 (± 1.00)**	4
BERT-MTL-LSTM	**88.136 (± 1.00)**	**84.64 (± 1.00)**	4
BERT-MTL-GRU	**89.47 (± 1.00)**	**86.19 (± 1.00)**	4

Table 3. Opinion Polarity Detection

(a) Fine-grained ABOM

Model	ACC	Macro F1	Epochs
MGAN [10]	81.49	71.48	-
BERT-PT [17]	84.95	76.96	10
BERT-LSTM [15]	82.21	72.52	10
BERT-Attention [15]	82.22	73.38	10
BAT [8]	82.27	73.7	10
DomBERT [18]	83.14	75.00	10
BERT-MTL-LSTM	**85.0**	**78.15**	4
BERT-MTL-GRU	**84.11**	**77.61**	4
BERT-MTL-Attention	**82.52**	**75.83**	4

(b) Coarse-Grained ABOM

Model	ACC	Macro F1	Epochs
GCN [19]	79.67	-	-
GCAE [19]	79.35	-	30
CapsNet-BERT [7]	86.55	-	-
AAL [21]	85.61	75.54	10
BERT-MTL-LSTM	**89.31**	**82.54**	4
BERT-MTL-GRU	**87.72**	**80.61**	4
BERT-MTL-Attention	**87.72**	**80.40**	4

6 Conclusions and Future Work

This paper proposes a novel BERT-based approach (BERT-MTL) for ABOM, which includes a Multi-Task learning model to extract the aspect terms and aspect categories simultaneously from the user reviews. Furthermore, we also perform Fine-grained and coarse-grained ABOM using the BERT model and applied three different pooling strategies on the last layer of the BERT to achieve high accuracy and take less amount of time to train. We achieved better results during the evaluation of the model than the previous approaches on the SemEval 2014 restaurant dataset. Also, some possible future works are: (a) performs soft parameter sharing between Fine-grained and coarse-grained ABOM models to build a more optimized model. (b) The adversarial or post-training in the MTL model can also increase the accuracy and F1-score.

References

1. Caruana, R.: Multitask learning. Mach. Learn. **28**(1), 41–75 (1997)
2. Cho, K., et al.: Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078 (2014)
3. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018)
4. Ejieh, C., Ezeife, C., Chaturvedi, R.: Mining product opinions with most frequent clusters of aspect terms. In: Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing (2019)
5. He, R., Lee, W.S., Ng, H.T., Dahlmeier, D.: An interactive multi-task learning network for end-to-end aspect-based sentiment analysis. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics (2019)
6. Hochreiter, S., Schmidhuber, J.: LSTM can solve hard long time lag problems. In: Advances in Neural Information Processing Systems, pp. 473–479 (1997)
7. Jiang, Q., Chen, L., Xu, R., Ao, X., Yang, M.: A challenge dataset and effective models for aspect-based sentiment analysis. In: Proceedings of the 2019 Conference on EMNLP-IJCNLP, pp. 6280–6285. Association for Computational Linguistics, Hong Kong, China, Novembers 2019
8. Karimi, A., Rossi, L., Prati, A., Full, K.: Adversarial training for aspect-based sentiment analysis with BERT. CoRR abs/2001.11316 (2020)
9. Kiritchenko, S., Zhu, X., Cherry, C., Mohammad, S.M.: Nrc-canada-2014: detecting aspects and sentiment in customer reviews. SemEval. **2014**, 437 (2014)
10. Li, Z., Wei, Y., Zhang, Y., Zhang, X., Li, X.: Exploiting coarse-to-fine task transfer for aspect-level sentiment classification. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, pp. 4253–4260 (2019)
11. Luong, M.T., Pham, H., Manning, C.D.: Effective approaches to attention-based neural machine translation. arXiv preprint arXiv:1508.04025 (2015)
12. Pontiki, M., Galanis, D., Pavlopoulos, J., Papageorgiou, H., Androutsopoulos, I., Manandhar, S.: Semeval-2014 task 4: aspect based sentiment analysis. Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), pp. 27–35 (2014). https://doi.org/10.3115/v1/S14-2004
13. Ruder, S.: An overview of multi-task learning in deep neural networks. arXiv preprint arXiv:1706.05098 (2017)
14. Schouten, K., Frasincar, F.: Survey on aspect-level sentiment analysis. IEEE Trans. Knowl. Data Eng. **28**(3), 813–830 (2015)
15. Sun, C., Huang, L., Qiu, X.: Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence. arXiv preprint arXiv:1903.09588 (2019)
16. Truşcă, M.M., Wassenberg, D., Frasincar, F., Dekker, R.: A hybrid approach for aspect-based sentiment analysis using deep contextual word embeddings and hierarchical attention. In: International Conference on Web Engineering. pp. 365–380. Springer (2020). https://doi.org/10.1007/978-3-030-50578-3
17. Xu, H., Liu, B., Shu, L., Yu, P.S.: BERT post-training for review reading comprehension and aspect-based sentiment analysis. arXiv preprint arXiv:1904.02232 (2019)
18. Xu, H., Liu, B., Shu, L., Yu, P.S.: Dombert: Domain-oriented language model for aspect-based sentiment analysis (2020)

19. Xue, W., Li, T.: Aspect based sentiment analysis with gated convolutional networks. arXiv preprint arXiv:1805.07043 (2018)
20. Xue, W., Zhou, W., Li, T., Wang, Q.: MTNA: a neural multi-task model for aspect category classification and aspect term extraction on restaurant reviews. IJCNLP **2017**, 151 (2017)
21. Zhu, P., Chen, Z., Zheng, H., Qian, T.: Aspect aware learning for aspect category sentiment analysis. ACM Trans. Knowl. Discov. Data **13**(6), 1–21 (2019)