2019 Seventh International Conference on Advanced Cloud and Big Data (CBD)

Multi-task Learning Based on Question-Answering Style Reviews for Aspect Category Classification and Aspect Term Extraction

Hanqian Wu[1,2], Zhike Wang[1,2], Mumu Liu[1,2], Jingyu Huang[3]
[1] *School of Computer Science and Engineering, Southeast University, Nanjing, China*
[2] *Key Laboratory of Computer Network and Information Integration of Ministry of Education, Southeast University, China*
[3] *State Grid HeBei Information Telecommunication Branch, China*
hanqian@seu.edu.cn, {zhikerwang, liudoublemu}@163.com, huangjy@he.sgcc.com.cn

***Abstract*—User-generated online reviews are booming in current Internet and e-commerce environment. The latest question-answering (Q&A)-style reviews are novel, easily digestible product reviews that also contain abundant referral information for customers. In this paper, we mine valuable aspect information of products contained in these reviews. To achieve this goal, we utilize two subtasks of aspect-based sentiment analysis: Aspect Term Extraction (ATE) and Aspect Category Classification (ACC). Most previous work focused on only one task or solved these two tasks separately, even though they are highly interrelated. To address this problem, we propose a novel multi-task neural learning framework to jointly handle these two tasks. We conducted extensive comparative experiments on annotated corpus and found that our proposed model outperforms several baseline models in ATE and ACC tasks, yielding significant strides in data mining for these types of reviews.**

I. Introduction

As social media and e-commerce platforms rapidly evolve, online product reviews are exceptionally popular. To process and analyze these textual reviews, Natural Language Processing (NLP) has garnered significant attention in recent years. Aspect-based sentiment analysis, as a important research topic in NLP field, offers fine-grained tasks for mining aspect information of product reviews. Accurately mining this information, however, involves three important subtasks: Aspect Term Extraction (ATE), Aspect Category Classification (ACC), and Aspect-level Sentiment Classification (ASC).

Recently, several studies have focused on ATE and ACC tasks, however, researchers regarded ATE and ACC as independent tasks and dealt with them separately, even though the tasks are highly interrelated. Intuitively, extracted aspect term information assists aspect category prediction, and aspect category information is advantageous to distinguish aspect terms from other words unrelated to aspect information. To overcome this problem, we propose a novel multi-task neural learning framework that jointly addresses the two tasks.

Question-answering (Q&A)-style reviews—a novel form of product reviews—consist of questions and answers where potential consumers generate questions and sellers or people who purchased the products provide answers. Fig. 1 shows an example of Q&A-style reviews with annotation information. Compared to conventional reviews, Q&A-style reviews effectively reduce the number of fake reviews, and make product information more credible. Thus, aspect-based sentiment analysis is particularly necessary and meaningful for mining valuable information contained in Q&A reviews.

Yet thus far, most studies do not focus on Q&A-style reviews because of the following difficulties. First, the corpus about Q&A-style reviews—especially the Chinese corpus—is scarce. However, we recently resolved this difficulty in previous work by designing a set of elaborate annotation rules and building a high-quality annotated corpus [2], [3]. Second, because of colloquial and informal nature of online reviews, existing word segmentation toolkits generate errors when dealing with text of Q&A-style reviews, which degrades the subsequent model's performance. As a solution, we adopt character-level rather than word-level embedding to represent Q&A text. Third, the ACC task for Q&A-style reviews is more difficult than for conventional reviews, because of the occasional irrelevant aspect terms. With this in mind, it only makes sense to focus on the aspect term mentioned in both the question and answer context. To this end, we leverage an attention mechanism to capture the most relevant aspect information contained in a Q&A text pairs.

Thus, by analyzing the effectiveness of multi-task model and characteristics of Q&A reviews, our research offers the following contributions:

- To contend with colloquial and informal nature of online reviews, we avoid word segmentation errors by adopting character-level rather than word-level embedding. In this way, our proposed model improves the the performance of ACC task and implements fine-grained extraction for ATE task.
- To address the occasional irrelevant aspect terms

978-1-7281-5141-0/19/$31.00 ©2019 IEEE
DOI 10.1109/CBD.2019.00056

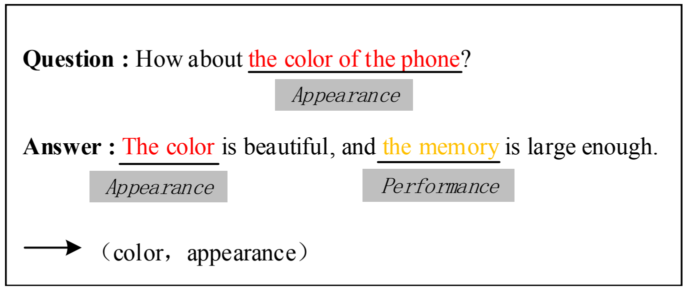

Fig. 1. Translated example of a question-answering (Q&A)-style review.

mentioned in the Q&A context, we introduce attention mechanism that captures the most relevant aspect information mentioned by both the question and answer contexts, which improves the performance of ACC task.

- To further improve the performance of ACC and ATE, we leverage the correlation between ATE and ACC to jointly address the two tasks.

II. Related Work

A. Aspect Category Classification

The ACC task can be treated as a supervised classification task. Given the predefined categories, our task is to identify a specified aspect term's category. Traditional approaches mainly focused on manually designing a set of features such as a bag-of-words or lexicon to train a classifier. Brychcin et al. [9] leveraged a set of binary Maximum Entropy (ME) classifiers for ACC. Kiritchenko et al. [8] used a set of binary Support Vector Machines (SVMs) with different types of n-grams and information from a specially designed lexicon. However, these approaches highly depend on the features' quality, and feature engineering is labor-intensive.

With the development of deep learning techniques, researchers have designed effective neural networks to address ACC task. Toh et al. [11] extracted features from words in every sentence and adopted the sigmoidal feedforward network to train a binary classifier. Xue et al. [14] proposed a multi-task neural network to jointly address ACC and ATE. (Our research differs from the work they built on conventional reviews, our proposed model is based on Q&A-style reviews and we leverage conditional random fields to further improve ATE's performance.) Wan et al. [16] proposed a semi-supervised method that combines neural networks with a logistic regression classifier for ACC.

B. Aspect Term Extraction

The ATE task extracts aspect and opinion terms explicitly contained in the sentence. Early work focused on researching rule-based methods. Hu and Liu et al. [7] leveraged frequent nouns or noun phrases to extract aspect terms, and tried to identify opinion terms by exploiting the relationships and occurrence between aspect terms and opinion terms. However, the rule-based approaches highly relied on hard-coded rules and external language resources. Later, ATE was treated as sequence tagging problem by using supervised featured-based methods such as Hidden Markov Models (HMMs) [1] or Conditional Random Fields (CRF) [5]. However, featured-based approaches greatly rely on features' quality—and again, feature engineering is both time-consuming and labor-intensive.

With the rapid development of neural networks, researchers proposed a neural language model for general high-level representations of words used to extract aspect terms. Liu et al. [6] used pretrained word embeddings as input of Recurrent Neural Network (RNN) for ATE. Yin et al. [15] proposed a hybrid method that first learns a distributed representation of words and dependency paths by RNN and then feeds the learned results along with some hand-crafted features into a CRF [4] for extracting aspect terms. Wang et al. [12] proposed a joint model consisting of Recursive Neural Network (ReNN) and CRF layer for ATE task. To reduce the influence of parsing errors, they further designed the RNN with coupled multilayer attention, to exploit the relationship of aspect terms and opinion terms for co-extraction [13].

III. Proposed Method

In this section, we describe the ATE and ACC tasks based on Q&A text pairs. On this basis, considering characteristics of Q&A-style reviews, we propose a multi-task model to jointly address the two tasks. Intuitively, the question text tends to be more important, because the aspect term needing categorization tends to appear in the question first. And then, the answer text also involves information related to the aspect term mentioned in the question context. Thus, we need to better model the representation of question text by doing a better job of harnessing relevant aspect information contained in both the question and answer context. Specifically, our proposed model uses two Bidirectional Long Short-Term Memories (Bi-LSTMs) to generate hidden state representations of the question and answer text, respectively. For the ATE task, we use a fully connected layer and CRF layer to extract the aspect term in the question text. For the ACC task, an attention mechanism is introduced to capture the most relevant aspect information between the Q&A text, and extend the representation of question text by leveraging the relevant aspect information contained in the answer text.

A. Aspect Term Extraction and Aspect Category Classification Tasks

We can treat the ATE task as sequence tagging problem, which extracts an explicit aspect term in the question text. Note that the extracted term could be a single word or a phrase. d **O** means others. In particular, if the aspect term is a single word, we label it as **S**. In this way, the From the sequence tagging perspective, the word tokens related to the given aspect category should be tagged according to a predefined label scheme. We define the label scheme as {**B**,**I**,**E**,**O**,**S**}, where **B** indicates an aspect term's beginning, **I** indicates the inside of an aspect term, **E** indicates an aspect term's end, anquestion text "How about the color of the phone?" can be tagged as "How/**O** about/**O** the/**B** color/**I** of/**I** the/**I** phone/**E** ?/**O**". Thus, we address the ATE task by training a sequence labeling model based on the combination of Bi-LSTM and CRF layers.

Instead of a sequence labeling model, the ACC task is approached as a general classification problem. Given the predefined categories, the task is to identify a aspect category for the specified aspect term. Thus, the proposed model uses two Bi-LSTMs to model representations of the question text and answer text, and then an attention mechanism is adopted to extend the representation of the question text for improving the ACC's performance.

B. Multi-task Model

Fig. 2 shows the architecture of multi-task learning framework. Given a Q&A-style review, assume that the question text $Q = \{w_1, w_2, ..., w_M\}$ contains M single words, where w_i represents the ith single word in the question text. Each single word is represented as $q_i \in R^{d_w}$ which is obtained from a word embedding matrix $E \in R^{d_w \times |V|}$, where d_w is the embedding dimension and |V| is the vocabulary size. Thus, we represent the question text as a character-level embedding matrix $S_Q = \{q_1, q_2, ..., q_M\}$. Similarly, we represent the answer text $A = \{s_1, s_2, ..., s_N\}$ as a character-level embedding matrix $S_A = \{a_1, a_2, ..., a_N\}$, where $a_j \in R^{d_w}$ denotes the jth single word of the answer text and N is the number of single words in an answer text.

Next, we feed the character-level embedding matrix of question text S_Q into a Bi-LSTM layer shared by the ATE and ACC tasks to generate a hidden state matrix of question text $H_Q = \{h_{q_1}, h_{q_2}, ..., h_{q_M}\}$, where we obtain the hidden state of each single word by averaging the forward and backward hidden state:

$$\overrightarrow{H_Q} = \overrightarrow{LSTM}(S_Q) \tag{1}$$

$$\overleftarrow{H_Q} = \overleftarrow{LSTM}(S_Q) \tag{2}$$

$$H_Q = AVG(\overrightarrow{H_Q}; \overleftarrow{H_Q}), \tag{3}$$

where $H_Q \in R^{d_h \times M}$, d_h is the dimension of hidden state and M is the number of single words in the question text.

1) Aspect term extraction: Given the hidden state matrix of question text H_Q, the model transforms it into an output label space using an additional fully connected layer:

$$P = H_Q^T \cdot W_{ate} + b_{ate}, \tag{4}$$

where $P \in R^{M \times N_t}$ is the output score matrix of N_t labels, and P_{ij} denotes the score of the jth tag of the ith single word in the question text. $W_{ate} \in R^{d_h \times N_t}$ and $b_{ate} \in R^{N_t}$ are parameters of the fully connected layer. Thus, according to the given question text $Q = \{w_1, w_2, w_3, .., w_M\}$, we define the prediction score for each output tag sequence $z = \{z_1, z_2, ...z_M\}$, where z_i denotes the label of w_i :

$$Score(Q, z) = \sum_{i=1}^{M-1} A_{z_i, z_{i+1}} + \sum_{i=1}^{M} P_{i, z_i}. \tag{5}$$

$A \in R^{N_t \times N_t}$ is a transition score matrix, and A_{ij} is the transition probability from label i to label j. Furthermore, we adopt a softmax function over all possible tag sequences for computing the posterior probability:

$$P(z|Q) = \frac{e^{Score(Q,z)}}{\sum_{\tilde{z} \in Z_Q} e^{Score(Q,z)}}, \tag{6}$$

where Z_Q denotes all possible tag sequence collections. According to the principle of maximizing posterior probability, we select the tag sequence that maximizes the posterior probability as the optimal path z^* and then extract aspect terms according to the tag sequence:

$$z^* = \arg\max_{\tilde{z} \in Z_Q} Score(Q, \tilde{z}). \tag{7}$$

2) Aspect category classification: Given the hidden state matrix of question text H_Q, the model uses another Bi-LSTM layer to obtain a hidden state matrix of answer text $A = \{s_1, s_2, ..., s_N\}$:

$$\overrightarrow{H_A} = \overrightarrow{LSTM}(S_A) \tag{8}$$

$$\overleftarrow{H_A} = \overleftarrow{LSTM}(S_A) \tag{9}$$

$$H_A = AVG(\overrightarrow{H_A}; \overleftarrow{H_A}), \tag{10}$$

where $H_A \in R^{d_h \times N}$ and N is the number of single words in the answer text. Noting that there may be irrelevant aspect terms mentioned in the Q&A text, we adopt an attention mechanism to capture the most relevant aspect information mentioned in both the question and answer context. Thus, the vector representation of question text

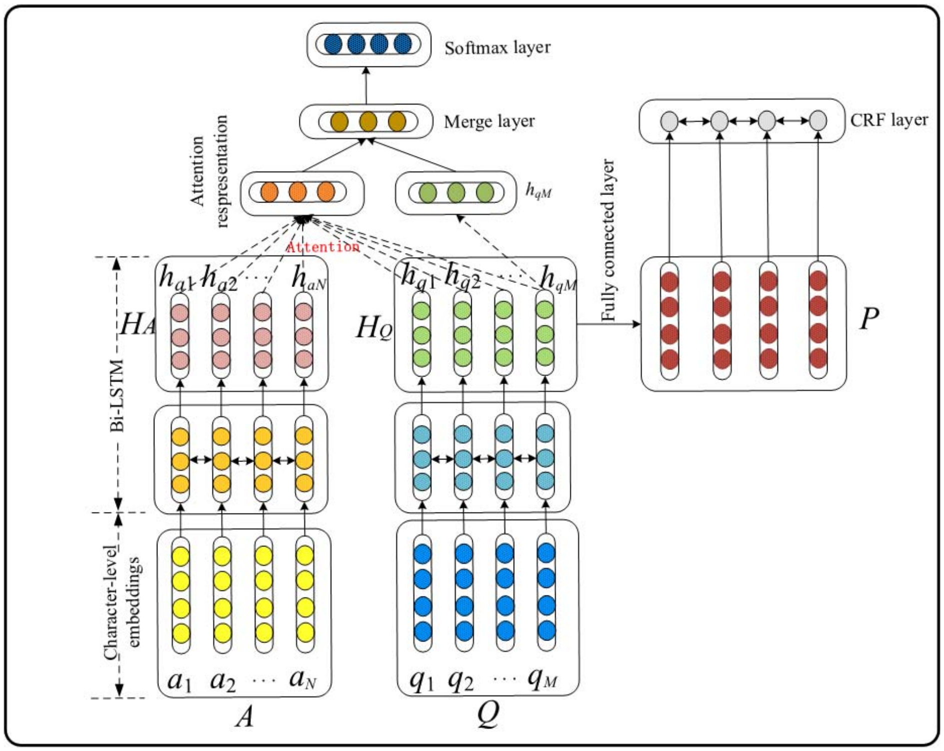

Fig. 2. Multi-task model architecture

could be enhanced by making full use of aspect information contained in the answer text. The attention layer calculates the attention representation of the question text according to the following formulas:

$$M = \tanh(W_c \cdot (H_A^T \cdot H_Q) + b_c) \tag{11}$$

$$\alpha = softmax(W_e^T \cdot M) \tag{12}$$

$$r = H_Q \cdot \alpha^T, \tag{13}$$

where $M \in R^{N\times M}$, r is the attention representation of question text, and $W_c \in R^{N\times N}$, $b_c \in R^M$, $W_e \in R^N$ are parameters to be trained. Finally, the final vector representation of question text is calculated by non-linearly combining r with the final hidden state h_{q_M} :

$$h^* = \tanh(W_f r + W_x h_{q_M}), \tag{14}$$

where $h^* \in R^{d_h}$, $W_f \in R^{d_h\times d_h}$ and $W_x \in R^{d_h\times d_h}$ are parameters to be trained. In the softmax layer, the final aspect category distribution of the given question text is predicted using the final vector representation of question text h^*:

$$y = softmax(Wh^* + b), \tag{15}$$

where $W \in R^{K\times d_h}$, $b \in R^K$ are parameters in the softmax layer and K is the number of predefined categories.

C. Model Training

In the ACC task, given a set of training data S_{Q_t}, S_{At}, and y_t, where S_{Q_t} is the tth question text, S_{At} is the tth answer text, and y_t is the ground-truth aspect category for the Q&A text pair (S_{Q_t}, S_{At}). We use the cross entropy function between y and y_t with L2 regulations as a loss function:

$$L_{acc} = -\sum_{t=1}^{N}\sum_{k=1}^{K} y_t^k \log y^k + \frac{l}{2} \parallel \theta \parallel^2, \tag{16}$$

where N is the size of training set, K is the number of predefined categories, l is the parameter for L2 regularization, and θ is a parameter set.

In the ATE task, given a set of training data S_{Q_t}, S_{At}, and z_t, where S_{Q_t} is the tth question text, S_{At} is the tth answer text, and z_t is the prediction-output tag sequence for the tth question text, assuming Score(S_{Q_t},z_t) is the score of tag sequence z_t, we describe the log-likelihood function as

$$L_{ate} = \sum_{t=1}^{N} Score(S_{Q_t}, z_t) - \log(\sum_{\tilde{z}\in Z_Q} e^{Score(Q,z)}), \tag{17}$$

where N is the size of the training set and Z_Q is all possible tag sequence collections. To learn the parameters of the multi-task model, we define the loss function as a weighted linear combination:

$$L = \lambda L_{acc} + (1-\lambda)L_{ate}, \tag{18}$$

where λ is the weight parameter. Parameters are optimized by using Adam optimization functions, and to solve overfitting problems, dropout strategy is adopted.

TABLE I
TRAINING DATA DISTRIBUTION

Data \ Domain	Digital	Beauty	Luggage
Aspect categories	7	10	11
The number of Q&A text pairs	2427	2927	2876
Most frequent aspect category	IO	Efficacy	quality
Q&A text pairs containing the most frequent aspect category	908	911	868
Maximum words of aspect term	8	8	7
Minimum words of aspect term	1	1	1

IV. Experiments

A. Experimental Setting

- **Data settings**: Our experiments use Q&A-style reviews as training data, which involves a digital domain, beauty domain, and luggage domain. Table 1 shows the distribution of experimental data. To avoid imbalanced data distribution, we discard the aspect categories that involve less than 50 Q&A text pairs.
- **Character-level representations**: As we mentioned, because of the informal nature of online reviews, we choose character-level rather than word-level embedding, which reduces the word segmentation errors. Specifically, we obtain character-level embedding by using Skip-gram [10] model provided by the gensim toolkit.
- **Evaluation metrics**: For the ACC task, the main evaluation metrics are Accuracy(A_{acc}) and F1-meature(F_{acc}), where F_{acc} is calculated as $F_{acc} = \frac{2P_{acc}R_{acc}}{P_{acc}+R_{acc}}$. For the ATE task, F_{ate} is calculated by the formula $F_{ate} = \frac{2P_{ate}R_{ate}}{P_{ate}+R_{ate}}$.
- **Hyperparameters**: We initialize all out-of-vocabulary words by sampling from the uniform distribution U(-0.01,0.01), the dimension of character-level embeddings, and hidden state vectors set to 300. We tune other hyperparameters according to development data, a model-use Adam optimizer with a batch size of 32, and an initial learning rate of 0.002. The weight parameters of the multi-task model λ is set to 0.55, and the dropout rate is set to 0.25 to reduce overfitting.

B. Experimental Results and Model Comparison

To comprehensively evaluate the performance of multi-task model, we compare our proposed model with several baselines for ATE and ACC based on Q&A reviews, respectively.

In the ACC task based on Q&A reviews, we compared the following baseline models:

- **LSTM(A)**: This model takes the answering text as input, and uses an LSTM network to model the answer text, then uses the hidden state representation of answer text for the ACC task.
- **LSTM(Q)**: This model takes the question text as input, and uses an LSTM network to model the question text, then uses the hidden state representation of question text for the ACC task.
- **LSTM(Q+A)**: This model takes question and answer text as input, and uses an LSTM network to model the question and answer text, then obtains the final hidden state representation by concatenating hidden state representations of question and answer text.
- **Bi-LSTM**: This model takes the question text as input, and uses the Bi-LSTM network to model the question text, then uses the hidden state representation for the ACC task.
- **Multi-task**: This model is a variation of our proposed model for ACC. Compared to ours, it ignores relevant information between question and answer text.
- **Multi-task+Attention**: This model is used for ACC task based on Q&A style reviews by constructing a multi-task learning framework. The ACC task is based on Bi-LSTMs, with attention mechanism is adopted to better model question text.

In the ATE task based on Q&A reviews, we compared the following baseline models:

- **CRF**: This method uses CRF to extract the aspect term from the question text. It uses character-level embedding learned from Skip-gram model as input.
- **Bi-LSTM**: This method uses Bi-LSTM to model question text, and then leverages a softmax layer for ATE task.
- **Bi-LSTM+CRF**: This method uses Bi-LSTM to model question text, and feeds hidden states into a CRF layer for ATE task.
- **Multi-task**: This model is a variation of our proposed model for ATE task. Compared to ours, it ignores the relevant information between Q&A text.
- **Multi-task+Attention**: This model is used for ATE task based on Q&A reviews by constructing a multi-task learning framework. The ATE task is conducted based on Bi-LSTM and CRF.

Tables II, III, and IV show the performance of proposed model with other baseline models in the digital,

TABLE II
RESULTS OF BASELINE METHODS IN THE DIGITAL DOMAIN

Models	Aspect Category Classification		Aspect Term Extraction
	A_{acc}	F_{acc}	F_{ate}
LSTM(Q)	0.779	0.668	-
LSTM(A)	0.663	0.466	-
LSTM(Q+A)	0.800	0.678	-
Bi-LSTM	0.833	0.764	0.586
CRF	-	-	0.573
Bi-LSTM+CRF	-	-	0.618
Multi-task	0.866	0.784	0.630
Multi-task+Attention(ours)	**0.887**	**0.839**	**0.636**

TABLE III
RESULTS OF BASELINE METHODS IN THE BEAUTY DOMAIN

	Aspect Category Classification		Aspect Term Extraction
	A_{acc}	F_{acc}	F_{ate}
LSTM(Q)	0.638	0.484	-
LSTM(A)	0.565	0.381	-
LSTM(Q+A)	0.662	0.536	-
Bi-LSTM	0.708	0.576	0.531
CRF	-	-	0.524
Bi-LSTM+CRF	-	-	0.539
Multi-task	0.724	0.609	0.547
Multi-task+Attention(ours)	**0.759**	**0.646**	**0.555**

TABLE IV
RESULTS OF BASELINE METHODS IN THE LUGGAGE DOMAIN

Models	Aspect Category Classification		Aspect Term Extraction
	A_{acc}	F_{acc}	F_{ate}
LSTM(Q)	0.638	0.513	-
LSTM(A)	0.469	0.346	-
LSTM(Q+A)	0.654	0.543	-
Bi-LSTM	0.688	0.626	0.523
CRF	-	-	0.507
Bi-LSTM+CRF	-	-	0.538
Multi-task	0.697	0.637	0.552
Multi-task+Attention(ours)	**0.725**	**0.655**	**0.576**

beauty, and luggage domains. By analysis, we draw the following conclusions.

For the ACC task, the performance of LSTM(Q) is obviously better than LSTM(A), which confirms that the question text tends to be more important than the answer text. Moreover, LSTM(Q+A) outperforms LSTM(Q) and LSTM(A), which indicates that combining the question and answer text could improve the performance of ACC task. The Multi-task model without an attention mechanism improves by 3.3%(A_{acc}) and 2.0%(F_{acc}) in the digital domain, 1.6%(A_{acc}) and 2.0%(F_{acc}) in the beauty domain, and 0.9%(A_{acc}) and 1.1%(F_{acc}) in the luggage domain compared to Bi-LSTM, which proves that the Multi-task model shows promise in improving ACC's performance with the help of extracted aspect information. Furthermore, the Multi-task+Attention model offers 2.1%(A_{acc}) and 5.5%(F_{acc}) improvement in the digital domain, 3.5%(A_{acc}) and 3.7%(F_{acc}) improvement in the beauty domain, and 2.8%(A_{acc}) and 1.8%(F_{acc}) improvement in the luggage domain compared to the Multi-task model. This confirms that the attention mechanism could capture the most relevant aspect information between the question and answer context, and improve the performace of ACC task by enhancing the question text's representation.

For the ATE task, the model Bi-LSTM+CRF improves by 3.2% in the digital domain, 0.8% in the beauty domain, and 1.5% in the luggage domain compared to Bi-LSTM, which proves that Bi-LSTM could learn the question text's context information, but the softmax layer ignores the interaction between tag sequences, and the CRF layer introduces a state transition matrix that makes use of sentence-level tag information to improve ATE's performance. The Multi-task model without an attention mechanism outperforms Bi-LSTM+CRF with an improvement of 1.2% in the digital domain, 0.8% in the beauty domain, and 1.4% in the luggage domain. This confirms that the aspect category information helps distinguish the aspect term from other words unrelated to aspect information. Further, the Multi-task+Attention model improves by 0.6%, 0.8%, and 2.4% in the domains compared to Multi-task model, which indicates that ACC's performance improvement further enhances ATE's performance.

Experimental results prove that our proposed multi-task model could make full use of the correlation between ACC and ATE to improve performance interactively. Also, the results confirm our two hypotheses about Q&A-style reviews: that the question text is more important than the answering text for the ACC task, and attention mechanism can capture the most relevant aspect information mentioned by both the question and answer context to improve ACC's performance.

C. Error Analysis

To determine our proposed model's limitations, we carefully analyzed the misclassified samples in the test set and found any factors that led to errors. The first factor is imbalanced data distribution, which makes the model tend to predict aspect categories that contain more Q&A text. For example, in the digital domain, 22.95 percent of misclassified samples are predicted to be "IO". Similarly, in the beauty and luggage domains, misclassified samples tend to predict "efficacy" and "quality." The second factor is that, in order to reduce word segmentation errors, we choose character-level rather than word-level embedding for ATE and ACC tasks. However, when we consider alternatives, modeling only one single word could disrupt our model's ability to model the real semantic and syntactic information of a clause or entire sentence, which would degrade the performance of subsequent model. The third factor is that the semantic information of some aspect terms are ambiguous in different contexts, which leads to difficulty in ACC task.

V. Conclusions and Future Work

In this paper, we propose a multi-task neural learning framework based on Q&A text for addressing ACC and ATE tasks simultaneously. The initial inspiration comes from our analysis of characteristics of Q&A-style reviews and the correlation between ACC and ATE tasks. Experimental results prove that our multi-task model outperforms other baseline models.

This work offers tremendous research value in helping consumers, reviewers, and companies with the burgeoning market of Q&A-style reviews. Although we made significant strides here and in previous work [2] [3], there are still many aspects that we hope to study, refine, and improve:

- To model clauses and entire questions better, we want to introduce a convolution operation over character-level embedding, and model local contextual information for improving ACC's performance.
- To overcome the training data's imbalanced distribution (because of the small scale of annotated corpus), we hope to use semi-supervised methods to improve performance.
- To consider and utilize the ways in which more than one relevant aspect terms are mentioned by both the question and answer contexts, we plan to conduct ACC for multiple aspect terms simultaneously.

Acknowledgments

This work is supported in part by the Industrial Prospective Project of the Jiangsu Technology Department under grant BE2017081 and the National Natural Science Foundation of China under grant 61572129.

References

[1] F. Li, C. Han, M. Huang, X. Zhu, Y.-J. Xia, S. Zhang, and H. Yu, "Structure-aware review mining and summarization," in Proceedings of the 23rd international conference on computational linguistics, 2010,pp. 653–661.

[2] Hanqian Wu, Mumu Liu, Jingjing Wang, Jue Xie, Chenlin Shen:Question-Answering Aspect Classification with Hierarchical Attention Network. CCL 2018: 225-237

[3] Hanqian Wu, Mumu Liu, Jingjing Wang, Jue Xie, Shoushan Li: Question-Answering Aspect Classification with Multi-attention Representation. CCIR 2018: 78-89

[4] John D. Lafferty, Andrew McCallum, Fernando C. N. Pereira:Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. ICML 2001: 282-289

[5] Shu, Lei, Xu, Hu, and Liu, Bing. Lifelong Learning CRF for Supervised Aspect Extraction[C]. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). 2017. 148–154.

[6] P. Liu, S. Joty, and H. Meng, "Fine-grained opinion mining with recurrent neural networks and word embeddings," in Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, 2015, pp. 1433–1443.

[7] M. Hu and B. Liu, "Mining and summarizing customer reviews," in Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ser. KDD '04. New York, NY, USA: ACM, 2004, pp. 168–177.

[8] S. Kiritchenko, X. Zhu, C. Cherry, and S. M. Mohammad, "NRCCananda-2014: Detecting aspects and sentiment in customer reviews," in Proc. 8th Int. Workshop Semantic Eval. (SemEval), Dublin, Ireland, 2014, pp. 437–442.

[9] T. Brychcin, M. Konkol, and J. Steinberger, "UWB: Machine learning approach to aspect-based sentiment analysis," in Proc. 8th Int. Workshop Semantic Eval. (SemEval), Dublin, Ireland, 2014, pp. 817–822.

[10] Tomas Mikolov, Ilya Sutskever, Kai Chen, Gregory S. Corrado, Jeffrey Dean:Distributed Representations of Words and Phrases and their Compositionality. NIPS 2013: 3111-3119

[11] Toh, Z., Su, J.: Nlangp: Supervised machine learning system for aspect category classification and opinion target extraction. In: International Workshop on Semantic Evaluation. pp. 496–501 (2015)

[12] W. Wang, S. J. Pan, D. Dahlmeier, and X. Xiao, "Recursive neural conditional random fields for aspect-based sentiment analysis," in Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, November 2016, pp. 616–626.

[13] W. Wang, S. J. Pan, D. Dahlmeier, and X. Xiao, "Coupled multi-layer attentions for co-extraction of aspect and opinion terms." in AAAI, 2017, pp. 3316–3322

[14] 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. In: Proceedings of the 8th International Joint Conference on Natural Language Processing. pp. 151–156. Asian Federation of Natural Language Processing (2017)

[15] Y. Yin, F. Wei, L. Dong, K. Xu, M. Zhang, and M. Zhou, "Unsupervised word and dependency path embeddings for aspect term extraction," arXiv preprint arXiv:1605.07843, 2016.

[16] Zhou, Xinjie, Wan, Xiaojun, and Xiao, Jianguo. Representation learning for aspect category detection in online reviews[C]. In: Twenty-Ninth AAAI Conference on Artificial Intelligence. 2015.