*Journal of Network Intelligence* ©2021 ISSN 2414-8105 (Online)
**Taiwan Ubiquitous Information** Volume **6**, Number **2**, May **2021**

# A joint aspect-based sentiment analysis method based on the encoder-decoder architecture

Jun-Ping Yao

Xi'an Research Inst. of High-Tech
ShaanXi Xi'an, 710025
junpingy200225@163.com

Kai-Yuan Cheng

Shaanxi Xi'an High-tech Research Institute
ShaanXi Xi'an, 710025
ckycky_com@163.com

Shou-Xiang Fan

Xi'an Research Inst. of High-Tech
ShaanXi Xi'an, 710025
fsx18840653260@163.com

Xiao-Jun Li

Xi'an Research Inst. of High-Tech
ShaanXi Xi'an, 710025
xi_anlxj@126.com

Received January 2021; revised March 2021

Abstract. *Aspect-based sentiment analysis serves as a basis for management of public opinions online and plays a crucial role in real-time detection of trending topics and shifts of sentiments on the Internet. Conventionally, aspect-based sentiment analysis is divided into two sub-tasks, which subjects the accuracy of aspect-based sentiment analysis to the result of aspect term extraction and makes it impossible to fully utilize the joint information from sub-tasks. Thus, in this study, we propose a multi-task sequence tagging feedback model based on the encoder-decoder architecture. To begin with,the aspect-based sentiment classification task is converted into a sequence sentiment tagging problem in the aspect term extraction model under the encoder-decoder architecture to synchronize aspect term extraction and sentiment classification tasks. Then,under the encoder-decoder architecture, the encoder encodes the text first and the decoder decodes the text word by word, so we integrate the tag of the preceding word and the correlations between the aspect and the sentiment of the current word into classification features to optimize the overall effect of aspect-based sentiment analysis. The experiment results show our proposed method achieved an F1 value that was 1.28%, 2.47% and 1.65% higher than two baseline methods on three datasets, which confirmed the effectiveness of our method in real-world application; meanwhile, our method performed better on the Restaurant datasets than on the Laptop dataset. This study is expected to provide a solution to optimizing the efficiency and accuracy of aspect-based sentiment analysis tasks. .*

**Keywords:**Sentiment analysis; Joint modelling; Correlation; Encoder-decoder.

1. **Introduction.** Usually, common sentiment classification tasks can be effectively performed using sentiment ontology [1] and word vector approaches. Different from ordinary sentiment classification, aspect-based sentiment analysis is a fine-grained classification task that requires analyzing the sentiment polarity (such as positive, negative, neutral) of different aspects of a sentence, playing a crucial role in opinion analysis, social networking, and false information monitoring. At present, aspect-based sentiment analysis methods divide the task into two separate sub-tasks — aspect term extraction and aspect-level sentiment classification, and most models built in previous works are intended for only one of the two sub-tasks. This conventional practice of dividing the task into two independent tasks, however, is problematic. The purpose of aspect-based sentiment analysis is two extract a two-tuple ( $g$, $s$ ) — an aspect term and its corresponding sentiment, from an unknown comment text; when the aspect-based sentiment classification task is divided into two separate sub-tasks, the final result is subject to the result of aspect term extraction, that is, the extracted aspect term is taken as the input for sentiment classification. In this scenario, only when the aspect term is correctly extracted can the sentiment be correctly classified, and consequently, the overall performance of aspect-based sentiment classification will be limited by the performance of aspect term extraction.

In recent years, in order to tackle this problem, some researchers have tried to integrate two independent sub-tasks into one task, and make use of the correlations between the two sub-tasks to improve their respective performance and thereby enhance the accuracy of aspect-based sentiment classification. For instance, Li et al. [2] proposed a unified model which adopts a unified tagging scheme and obtains aspect and sentiment features through one-time sequence tagging. He et al. [3] proposed an interactive multi-task learning network (IMN), which not only synchronizes aspect term extraction and sentiment analysis, but introduces the document-level classification task to optimize the shared underlying parameters and realize multi-task aspect-based sentiment analysis. The major challenge in building these models is to construct and utilize the correlations between the two sub-tasks to improve the overall performance of aspect-based sentiment analysis. Although the two fusion models mentioned above link the two subtasks together, both models have certain drawbacks. The model proposed by Li et al. [2] constructs the correlation through unified tagging and is short of an explicit message passing mechanism. In the model proposed by He et al. [3], the aspect-level sentiment classification task receives the sentiment features output from the aspect term extraction task, which in fact increases the risk of error transfer. Also, it is believed that when the text has complex semantic connotations, the sentiment of the text is not conveyed by words of clear sentimental polarities, but is hidden in the semantic context of the text or part of the text.

In this logic, we propose an aspect-based sentiment analysis method based on the encoder-decoder architecture. The method integrates the sub-tasks — aspect term extraction and aspect-level sentiment classification, and realizes unified extraction of opinions. Based on a sequence tagging feedback model for aspect term extraction proposed by Fan et al. [4], we migrate the aspect-level sentiment classification model onto an encoder-decoder architecture of the aspect term extraction model, convert the problem of aspect-level sentiment classification into a task of decoding and tagging the sentimental features of the current word that is performed simultaneously with the aspect information extraction task, thereby synchronizing aspect term extraction and sentiment analysis. Meanwhile, when tagging the aspect and sentiment features of the current word, the encoder-decoder architecture is utilized to integrate the tag of the preceding word and the correlations between the aspect and sentiment features of the current word into the classification features to increase the utilization rate of the inter-task correlations and optimize the overall performance of aspect-based sentiment analysis.

2. **Research status quo in China and abroad.** Aspect-based sentiment analysis is fine-graine sentiment analysis that extracts opinions from comments. Most aspect-based sentiment analysis research divides the task into two sub-tasks, i.e., aspect term extraction [5–9] and aspect-level sentiment classification [10–14], and the majority of these studies are focused on one of the two sub-tasks. This is because the two sub-tasks differ in implementation: aspect term extraction is a sequence tagging task, while aspect-level sentiment classification is a classification task based on given aspect terms. Aspect term extraction can be regarded as a phrase-level classification task, and aspect-level sentiment classification can be regarded as a sequence tagging task; these two tasks are correlated and mutually contributary. For instance, the aspect and sentiment tags are dependent, and the models share parameters. Some scholars have adopted the joint modelling approach for these two tasks, and as sequence-tagging the implicit aspect terms is unattainable, most of previous works are devoted to explicit aspect-based sentiment analysis.

Most joint modelling methods principally follow the aforementioned framework — using different but explainable combinations to extract aspect terms and corresponding sentiment orientations simultaneously. These methods generally follow two technical routes: (1) building a joint model in which two sub-tasks are constructed and the inter-task correlations are established, and the sum loss of the two sub-task modules is taken as the overall loss to update the parameters. The IMN model built by He et al. [3] is a typical joint model. In their model, the document-level classification task is merged into the aspect extraction and sentiment analysis tasks, and the aspect classification probability distribution result of the aspect extraction task is used to compute the self-attention weights of sentiment classification; they also designed an message passing mechanism that aggregates the outputs of four tasks and use the knowledge to update the shared latent vectors, which not only utilizes the inter-task correlations but benefits from the outcomes of multi-task learning. Luo et al. [5] used two recurrent neural networks to extract aspect and sentiment vectors, and input these vectors into a shared unit to extract correlations before sending the correlation information into the two tasks. (2) Building a unified model that unifies two sub-tasks into one; in these models, the boundaries between tasks are blurred, and a unified tagging scheme is established. For instance, B-POS, NEG, NEU, I-POS, NEG, NEU, E-POS, NEG, NEU, S-POS, NEG, NEU [2] represent the sentiment orientations (polarities) of the beginning word, the intermediate word, the ending word and the single word, and the other words are tagged as O. This unified tagging method can distinguish the boundaries of the aspect and the sentiment. For example, Zhang et al. [16] and Mitchell et al. [17] established unified tagging schemes, used linguistic features as inputs into the conditional random field (CRF) model to judge the aspect boundaries and sentiment orientations. Li et al. [2] used a double-layered bi-LSTM network to extract aspect terms and sentiment features simultaneously, and output unified tags by way of three modules responsible for boundaries, sentiment and target detection, respectively.

Many researchers have compared joint models and unified models with pipeline models that divide aspect-based sentiment analysis into separate sub-tasks, and the results revealed that none remained the optimal choice for all scenarios, and the model should be chosen as per the specifics of the task. For instance, in the experiments by Mitchell et al. [17], the unified model outperforms other models when the data size is small, but as the data size increases, the performance of the pipeline model and the joint model improves significantly. Luo et al. [15], however, hold that the joint model is more flexible than other two, and the unified model is likely to mess up the learnt features because it combines the sentiment tags with the aspect tags. Hu et al. [18] argue that the traditional tagging schemes have a large search space, and thus propose an end-tagging scheme. They tested their scheme on the three types of models, and found that the pipeline model showed the best performance under their tagging scheme.

3. **Multi-task sequence tagging feedback model based on the encoder-decoder architecture.** This study proposes an aspect-based sentiment analysis model based on the encoder-decoder architecture. The model is an extension of the model put forward by Fan et al. [4], and it is a joint model intended to improve the accuracy of aspect-based sentiment analysis. Specifically, we convert the aspect-based sentiment classification model from a model based on the aspect terms to a model that is based on the sequence sentiment tag of the current word, migrate the model under an encoder-decoder architecture, establish the correlations between the current word and the preceding word and inter-task correlations, to improve the overall performance of aspect-based sentiment analysis. As the two sub-task modules rely on our previous model [4], we are not going to belabor the work and will focus on the adjustments we have made in this study. The new model will be introduced in detail in terms of the model migration and adjustment, correlation establishment, model loss and training. Figure 1 shows the structure of our new model.

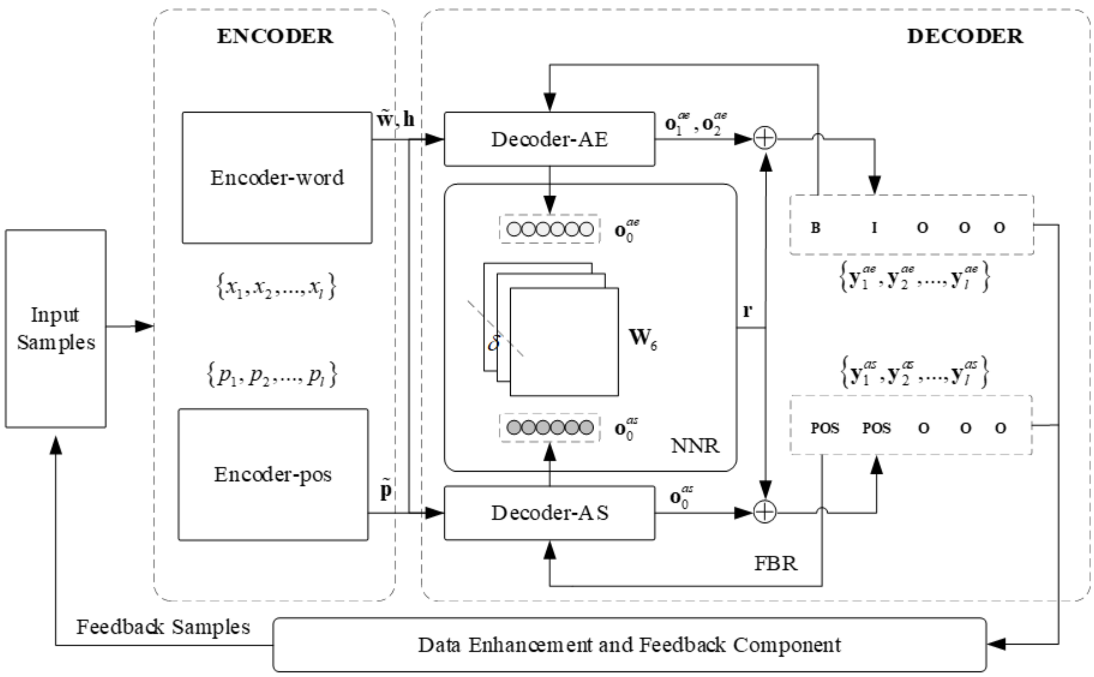


FIGURE 1. Structure of our proposed joint model

3.1. **Description of problems and interpretation of symbols.** A multi-task joint model for aspect-based sentiment analysis is built. With a given text sequence $S = \{x_1, x_2, ..., x_l\}$ that contains given aspect terms and corresponding sentiment representations, the model extracts the aspect terms $A = \{x_{l-i}, x_{l-i+1}, ..., x_{l-i+k}\}(k \leqslant i)$ , and meanwhile proposes the question concerning the sentiment or polarity ($C$) of A: whether it is positive, negative or neutral. Tagging of the aspect terms adopts a three-level tagging scheme {B,I,O} , and the aspect-level sentiment classification adopts a four-level tagging scheme {POS,NEG,NEU,O}.For the convenience of model description, we use $P = \{p_1, p_2, ..., p_l\}$ to represent the part-of-speech sequence of the text sequence, $\mathbf{w} = \{\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_l\}$ to represent the corresponding vector sequence, $\mathbf{p} = \{\mathbf{p}_1, \mathbf{p}_2, ..., \mathbf{p}_l\}$ to represent the corresponding part-of-speech vector sequence, and $\hat{\mathbf{y}}^m = \{\hat{\mathbf{y}}_1^m, \hat{\mathbf{y}}_2^m, ..., \hat{\mathbf{y}}_l^m\}$

to represent the one-hot encoding vector matrix of aspect and sentiment tags, where $m \in \{ae, as\}$ .

3.2. **Model migration and adjustment.** As Figure 1 shows, the integrated model consists of four modules, i.e., an encoding module, an aspect-level sentiment classification decoding module, a data enhancement module and a feedback module. To merge two separate models into a joint model entails three steps: initial encoding integration, migration and adjustment of the aspect-level sentiment classification model, and adjustment of the data enhancement component.

3.2.1. *Initial encoding integration.* Though both the two sub-tasks of aspect term extraction and aspect-level sentiment classification employ the double-layer bi- Gated Recurrent Unit (GRU) to encode the initial text vector matrix, the difference is that in the sub-task of aspect extraction extraction, the part-of-speech vector matrix is separately encoded to differentiate the roles of part-of-speech information under different language expression modes during the decoding process; while in the aspect-level sentiment classification sub-task, the word vector and the part-of-speech vector are merged before being input the double-layer bi-directional GRU for encoding; to maximize integration of the shared underlying encoding network, the aspect-level sentiment classification module is migrated to the encoder-decoder architecture of the aspect term extraction module, the output joint matrix obtained through two-way encoding by the aspect extraction task encoder is transformed into inputs into the sentiment extraction module in the aspect-based sentiment classification model. The specific calculation is shown in Eqs. (1) – (3).

$$\tilde{\mathbf{w}} = \mathrm{Bi}-GRU(\mathbf{w})\mathbf{W}_1 + \mathbf{b}_1 \tag{1}$$

$$\tilde{\mathbf{p}} = \mathrm{Bi}-GRU(\mathbf{p})\mathbf{W}_2 + \mathbf{b}_2 \tag{2}$$

$$\mathbf{O} = \tanh([\tilde{\mathbf{w}}; \tilde{\mathbf{p}}]\mathbf{W}_3 + \mathbf{b}_3) \tag{3}$$

where $\tilde{\mathbf{w}} \sim \tilde{\mathbf{p}} \in R^{n\times l}$ represent the vector matrices for the word vector matrix and the part-of-speech matrix of the text sequence output by the encoder, n is the feature dimension, l is the length of the current text sequence, $\mathbf{o} \in R^{n\times l}$ is the input vector matrix of three types of sentiments (positive, negative and neutral) of the sentiment classification module,$\mathbf{W}_1 \sim \mathbf{W}_3$ , and $\mathbf{b}_1 \sim \mathbf{b}_3$ are learnable parameters.

3.2.2. *Migration and adjustment of the aspect-level sentiment classification model.* As the aspect term extraction model is a sequence tagging model based on the encoder-decoder architecture, and in the decoding process, three types of features, i.e., target word semantic features, part-of-speech features, and dependent features, are used for word-by-word decoding and classification; the aspect-level sentiment classification model is a classifier based on the aspect of the current text, that is, when a text sequence S and its aspect A are given, the model extracts the overall sentiment features of the A in the context of S. Thus, to migrate the aspect-level sentiment classification model into the aspect term extraction model to build a multi-task joint model, we need first to convert the aspect-level sentiment classification model into a sequence tagging model. As the boundaries of three types of sentiments are defined based on the current aspect, when the joint model is built, the aspect becomes the predictive features which cannot be used as inputs into the aspect-level sentiment classification model. Therefore, in this study, we propose converting the aspect-level sentiment classification model into a sequence sentiment tagging model based on the current word, that is, when computing and extracting the three types of sentiment features, the model replaces the encoding features of aspects by the encoding features of the current word. The aspect-based sentiment classification model in Eqs. (1) –(3) is adjusted into the following:

$$\mathbf{s} = \text{Attention}(\mathbf{O}, \tilde{\mathbf{w}}_t)/\, \mathrm{d}(S, x_t) \tag{4}$$

$$\mathbf{c} = \mathbf{O}\mathbf{s} \tag{5}$$

$$\tilde{\mathbf{w}}_t = \mathbf{c} + \tilde{\mathbf{w}}_t \tag{6}$$

where $\mathbf{O} \in R^{n\times l}$ stands for the output vector matrix of the text sequence after transformation, $\tilde{\mathbf{w}}_t \in R^n$ is the feature vector of the current word with the sentiment features combined; the initial value is the encoded output vector of the current word, $\mathrm{d}(S, x_t)$ is the attention attenuation coefficient obtained with the current word as the benchmark, $\mathbf{s} \in R^l$ is the value of attention given by the current aspect to each word in the text sequence, $\mathbf{c} \in R^n$ is the sentiment feature vector of the current aspect . Meanwhile, to convert the task of aspect-based sentiment extraction into a current word-based sentiment extraction task, we do not need to recalculate different aspect feature representations of multiple words, and thus the Eqs. (4) and (6) can be dispensed with in the original aspect-based sentiment classification model.

3.2.3. *Adjustment of the data enhancement component.* In the data enhancement and feedback component, word deletion is performed on misclassified samples to generate new samples that will then be fed back to the sample set. In the aspect-level sentiment classification module, however, the sequence should have a certain length for extraction of multi-point sentiment features, and if the sequence is too short, problems like lack of sentiment features and decreased distinction between sentiment features will emerge. To address these problems, we propose the method of increasing the text length recognition and control parameter during the word deletion operation. Specifically, when the length of the misclassified sample is smaller than , the model will exchange and convert the words instead of deleting the word.

3.3. **Establishing correlations.** Fully utilizing correlations is the key to improving the overall performance of aspect-based sentiment analysis after integrating the two subtasks of aspect extraction and aspect sentiment classification into a multi-task joint model. Observations show strong correlations between the tags of the current word and the preceding word as well as between the aspect and sentiment tags of the current word, as shown in Table 1. The correlations can be divided into two categories — one is the sequence tagging correlation, and the other is inter-task tagging correlation. To obtain the former type of correlation, we can introduce the preceding-word tag into the current-word tagging features in a text sequence to construct the classification features of the current word. This applies principally to the cases in which the tags between multiple words in a sequence are dependent. For instance, the sentiment tags for words in a multi-word aspect term are the same. The inter-task tagging correlation mainly applies to improving the accuracy and synchronicity of initial judgement of the aspect and sentiment tags. Specifically, the correlations between the classification features of the current aspect and the aspect-level sentiment are computed to obtain the correlation vectors, which are then added to the respective classification features so that the accuracy and synchronicity of the respective classification sub-task can be improved. As the aspect extraction module in the joint model adopts the two-way decoding scheme, the model needs to obtain first-level overall classification features based on the first-level classification features of both ways and then achieve the correlation vectors in combination with the aspect-level sentiment classification features. Eqs.(7) - (11) show the calculation process of the correlation vectors.

TABLE 1. Correlations between aspect and sentiment tags.

| No. | Condition | | | Correlation | |
|---|---|---|---|---|---|
| | Preceding-word aspect | Preceding-word sentiment | Current-word aspect | Current-word aspect | Current-word sentiment |
| 1 | B | POS | | Not B | POS or O |
| 2 | | NEG | | Not B | NEG or O |
| 3 | | NEU | | Not B | NEU or O |
| 4 | I | POS | | Not B | POS or O |
| 5 | | NEG | | Not B | NEG or O |
| 6 | | NEU | | Not B | NEU or O |
| 7 | O | O | | Not I | - |
| 8 | | | B | | Not O |
| 9 | | | I | | Not O |
| 10 | | | O | | O |

$$\mathbf{o}_1^{ae} = \mathbf{g}_{t1} \times \tilde{\mathbf{w}}_t + \mathbf{g}_{t2} \times \mathbf{o}_t^{ae} \tag{7}$$

$$\mathbf{o}_2^{ae} = \mathbf{g'}_{t1} \times \tilde{\mathbf{w}}_t + \mathbf{g'}_{t2} \times \tilde{\mathbf{o}}_t^{ae} + \mathbf{g'}_{t3} \times \tilde{\mathbf{p}}_t^{e} \tag{8}$$

$$\mathbf{o}_0^{ae} = \mathbf{W}_4[\mathbf{o}_1^{ae}; \mathbf{o}_2^{ae}] + \mathbf{b}_4 \tag{9}$$

$$\mathbf{o}_0^{as} = \mathbf{W}_5[\mathbf{q}; \mathbf{h}_\eta; \mathbf{h}_\lambda; \tilde{\mathbf{y}}_{t-1}^{as}] + \mathbf{b}_5 \tag{10}$$

$$\mathbf{r} = \tanh((\mathbf{o}_0^{as})^T \mathbf{W}_6 \mathbf{o}_0^{ae}) \tag{11}$$

where $\mathbf{o}_1^{ae}, \mathbf{o}_2^{ae} \in R^n$ are the two-way first-level classification features in the aspect extraction module. As per the equations for the original aspect extraction model, i.e., Eqs. (7) and (8), $\mathbf{o}_0^{ae} \in R^n$ is the overall first-level aspect classification feature vector of the target word, $\mathbf{o}_0^{as} \in R^n$ is the sentiment classification feature vector of the target word, into which the preceding word sentiment tag vector $\tilde{\mathbf{y}}_{t-1}^{as} \in R^\kappa$ is added to enhance the continuity of judgement of the aspect sentiment. As with the the aspect extraction model, the preceding-word tag in the training process is the actual tag of the preceding word, while in the testing process, the tag is the predictive tag; the tagged word embeddings adopt random initialization, and $\kappa$ represents the dimension of the sentiment tag vector; $\mathbf{r} \in R^\delta$ is the correlation vector between the aspect and sentiment of the target word, $\delta$ is the correlation vector dimension, which is obtained by the three-dimensional tensor — $\mathbf{W}_6 \in R^{\delta \times n \times n}$ . The three-dimensional tensor is a set of multiple correlation matrices, and each correlation matrix is multiplied by the $\mathbf{o}_0^{ae}$ and $\mathbf{o}_0^{as}$ to obtain a specific correlation feature; and by combining multiple correlation features, we can obtain the correlation vector between $\mathbf{o}_0^{ae}$ and $\mathbf{o}_0^{as}$ . By fusing the correlation vectors into the classification feature vectors of each module, we can output the two-way first-level classification result and the final classification feature vector of aspect-level sentiment, as shown in the following equations.

$$\mathbf{r}_t = \mathbf{W}_7[\mathbf{o}_1^{ae}; \mathbf{r}] + \mathbf{b}_6 \tag{12}$$

$$\mathbf{r'}_t = \mathbf{W}_8[\mathbf{o}_2^{ae}; \mathbf{r}] + \mathbf{b}_7 \tag{13}$$

$$\mathbf{o}^{as} = [\mathbf{o}_0^{as}; \mathbf{r}] \tag{14}$$

Substituting $\mathbf{r}_t$ and $\mathbf{r'}_t$ into Eq. (9) and $\mathbf{o}^{as}$ into Eq. (15), we obtain the probability distribution of target-word aspect and sentiment classes.

3.4. **Model loss and training.** The cross entropy is used to measure the loss of the aspect extraction model and the aspect-level sentiment classification model; and the sum loss of the two sub-models is the total loss of the joint model. Calculation of the total loss is as follows.

$$L(\theta_{ae}, \theta_{as}) = L_1(\theta_{ae}) + L_2(\theta_{as}) = -\frac{1}{N}\sum_{i=1}^{N}\frac{1}{l_i}\sum_{j=1}^{l_i}(\hat{\mathbf{y}}_{i,j}^{ae}\log(\mathbf{y}_{i,j}^{ae})^T + \hat{\mathbf{y}}_{i,j}^{as}\log(\mathbf{y}_{i,j}^{as})^T) \quad (15)$$

where $\mathbf{y}_{i,j}^{ae} \in R^{C_{ae}}$ and $\mathbf{y}_{i,j}^{as} \in R^{C_{as}}$ represent the aspect and sentiment probability distribution of the $j^{th}$ word in the $i_{th}$ sample; $\hat{\mathbf{y}}_{i,j}^{ae} \in R^{C_{ae}}$ and $\hat{\mathbf{y}}_{i,j}^{as} \in R^{C_{as}}$ represent the one-hot encoding vectors of the real aspect and sentiment classification labels of the $j^{th}$ word in the $i_{th}$ sample, N is the number of training samples, $l_i$ is the is the number of words in the current training sample, $\theta_{ae}$ and $\theta_{as}$ represent all parameters in the aspect extraction module and the aspect-level sentiment classification module.

## 4. Experiment.

4.1. **Experiment data and setting.** This study used three datasets —the Laptop comment dataset and Restaurant comment dataset in SemEval-2014 Task 4 [19] and the Restaurant comment dataset in SemEval-2015 Task 12 [20], to train and test our proposed joint model, and the comment samples with the tag "Conflict" were tagged as samples without aspects. Table 2 shows the specifics of the datasets. To facilitate horizontal comparison between our proposed model with other methods, we used the universal Glove.840B.300d pretrained word vectors with a dimension of 300; the part-of-speech vectors, aspect tag vectors and sentiment tag vectors were randomly initialized by uniform distribution $u \sim (-1/\sqrt{d}, 1/\sqrt{d})$ , where d is the vector dimension; the part-of-speech vector dimension was set at 300 ($n = 300$), the dimension of the aspect tag vector and the sentiment tag vector was set at 50 ($\kappa = 50$), and the correlation vector dimension was set at 100 ($\delta = 100$). The NLTK [21] (Natural Language Toolkit) was used to segment the document into word sequences, and Stanford-CoreNLP [22], a natural language toolkit developed by Stanford University, was employed to decipher the part of speech of the comment texts to obtain the part-of-speech information of words. To increase the model's efficiency, we pretrained and converted the comment data into indexed phrase sequences and indexed part-of-speech sequences, and all phrases involved in the comments were extracted as numpy arrays from the original complete word vectors, while out-of-vocabulary words were randomly initialized to increase the loading speed of word vectors. The tags of training samples were converted into one-hot encoding matrices according to the number of tags, and the 10% of the training samples were randomly selected to serve as the test dataset.

Table 2. Specifics of the sample data.

| Dataset | | Number of comments | Number of aspects | Positive | Negative | Neutral |
|---|---|---|---|---|---|---|
| Laptop14 | Train | 3045 | 2328 | 994 | 870 | 464 |
| | Test | 800 | 638 | 341 | 128 | 169 |
| Restaurant14 | Train | 3044 | 3608 | 2164 | 807 | 637 |
| | Test | 800 | 1120 | 728 | 196 | 196 |
| Restaurant15 | Train | 1315 | 1188 | 902 | 252 | 34 |
| | Test | 685 | 525 | 319 | 179 | 27 |

The size of all hidden layers of the GRU network in the model was set at 300, and the dropout strategy was employed to avoid parameter overfitting, with a dropout rate set at 0.1; the batch size was set at 16, and the Adam [23] optimization algorithm was employed to upgrade the model parameters, with a learning rate set at 0.001 and an epoch of 20. As for evaluation indicators, the F1 value was employed to evaluate the aspect tags, marked as A-F1, the accuracy and F1 value were employed to evaluate the aspect-level sentiment tags, marked as S-ACC and S-F1, and the F1 value was used to evaluate the overall classification performance of the joint model, marked as I-F1.

It should be noted that in the training process, the calculation of the loss of the aspect-level sentiment classification module takes into account the predictive information of all phrases in the sequence to improve the model's capacity to understand the sentiment expression mode; in the testing process, measurement of the accuracy of the sentiment tags considers only the sentiment tags within the boundary of the actual aspect; because sentiment tagging is susceptible to the problem of discontinuity, for multi-word aspect extraction, if the sentiments of the majority of phrases within the boundary are tagged correctly, we consider the sentiment tagging is correct; when the numbers of words of different sentiments are equal and the sentiment of the phrase cannot be determined, the aspect-level sentiment of the first word is considered as the current aspect-level sentiment. As for the indicator I-F1 that indicates the overall classification performance, only when both the aspect and the corresponding sentiment are tagged correctly can we consider that the sentiment is classified correctly.

4.2. **Comparison Experiments.** In the experiments, we compared our proposed model with three types of models. The first is the pipeline model. In the joint model, the aspect-level sentiment classification model was converted into a current word-based sequence tag sentiment classification model, which changed the structure of the original model, so in this study, we built two pipeline models: Pipeline1 is the aforementioned model which predicts the aspect and the sentiment separately in a pipeline manner; in Pipeline2, the aspect extraction module remains unchanged, but the aspect-level sentiment classification module is displaced by the sequence tag sentiment classification module in the joint model. By comparing our model with these two pipeline models, we confirmed that our proposed joint model improved the overall performance of sentiment extraction. The second type is the joint model. We compared our model with the IMN model proposed by He et al. [3], and confirmed the validity of our proposed model. The third type is the unified model. We compared our model with the Uni-ABSA model proposed by Li et al. [2] in terms of their performance in extracting opinions under different tagging systems.

4.3. **Experiment results and analysis.** Table 3 shows the comparison between our proposed model and other types of models. As the table shows, our proposed model demonstrates good performance on three datasets. When compared with the Pipeline1 model on the Laptop14 dataset, our model shows no superior performance and even achieves a lower I-F1 value, but it performs better in extraction of aspect and sentiment features; on the Restaurant14 and Restaurant15 datasets, our model performs better, with an I-F1 value 1.51% and 1.20% higher than the Pipeline1 model, which indicates that the aspect-level sentiment features in the Restaurant comments are easier to capture. Likewise, our model performs significantly better than Pipeline2 model on the Restaurant14 and Restaurant15 datasets, achieving an I-F1 value 0.2%, 2.46% and 2.47% higher than the latter model on the three datasets, respectively. It indicates that under our proposed architecture, the correlations established in our study play a positive role in improving the performance of aspect extraction, aspect-level sentiment classification and the overall sentiment analysis, and this role is more prominent on Restaurant comments than on the Laptop comments. Comparison between Pipeline1 and Pipeline2 reveals that when the aspect-level sentiment classification module is converted into a sequence tag sentiment classification module, the model's performance on sentiment classification and the overall sentiment analysis decreases, which indicates the aspect information has salient impacts on the sentiment classification task, and using the current word as the benchmark for sentiment extraction reduces the weight of the aspect feature in sentiment feature extraction.

TABLE 3. Performance of different models on different datasets

| Dataset | Indicator | PIPLINE1 | PIPLINE2 | IMN* | Uni-ABSA* | OURS |
|---|---|---|---|---|---|---|
| Laptop14 dataset | AE-F1 | 81.52 | 81.52 | 76.96 | 77.34 | **81.85** |
| | AS-Acc | 74.35 | 73.05 | 72.89 | 72.30 | **74.79** |
| | AS-F1 | 70.68 | 69.87 | 67.26 | 68.24 | **71.21** |
| | I-F1 | **57.62** | 57.12 | 56.25 | 55.88 | 57.34 |
| Restaurant14 dataset | AE-F1 | 83.80 | 83.80 | **83.95** | 83.92 | 83.72 |
| | AS-Acc | 80.55 | 79.34 | 79.65 | 79.68 | **80.63** |
| | AS-F1 | 72.91 | 71.88 | 69.32 | 68.38 | **73.04** |
| | I-F1 | 67.74 | 66.79 | 66.96 | 66.60 | **69.25** |
| Restaurant15 dataset | AE-F1 | **70.35** | **70.35** | 69.23 | 69.40 | 70.15 |
| | AS-Acc | 82.23 | 82.42 | 81.64 | 82.56 | **83.66** |
| | AS-F1 | 58.37 | 57.01 | 57.51 | **58.81** | 58.57 |
| | I-F1 | 57.44 | 56.27 | 56.80 | 57.38 | **58.74** |

Note:
1.Statistic for models marked with "*" are from He et al's *work*[3]; IMN is termed as "$IMN^{-d}$ wo DE" in He et al.'s work;bolde numbers are the highest value of the corres- -ponding indicator.
2.The test results were obtained under the length control parameter condition $\eta = 6$ and $\xi = 10$.
3.The test data on the Laptop14 dataset are obtained when $\alpha = \beta = \gamma = 0.3$;the test data on Restaurant14 and Restaurant15 datasets when $\alpha = 0.3, \beta = 0.4$ and $\gamma = 0.5$.

4.4. **Ablation experiments.** We also compared our model with the model with the model with the current-word aspect and sentiment correlations removed (OURS-w/o-NNR) and the model with the correlations between the sentiment of preceding and current words removed (OURS-w/o-FBR) to test the impacts of these two types of correlations on the joint model's performance. Table 4 shows the experiment results. As the results show, both of these two types of correlations play an important role in improving the model's performance, but the correlation between the aspect and sentiment of the current word plays a more salient role than the sentimental correlations between the preceding word and the current word. That's because the correlations between the preceding word and the current word are focused on the internal sentiment correlations of the aspect terms and weighs less in the aspect-level sentiment analysis task.

TABLE 4. Specifics of the sample data.

| Dataset | Indicator | OURS | OURS-w/o-NNR | OURS-w/o-FBR |
|---|---|---|---|---|
| Laptop14 | AE-F1 | **81.85** | 81.36 | 80.27 |
| | AS-Acc | **74.79** | 73.47 | 73.66 |
| | AS-F1 | **71.21** | 70.84 | 70.42 |
| | I-F1 | **57.34** | 56.85 | 56.73 |
| Restaurant14 | AE-F1 | 83.72 | 83.51 | 83.61 |
| | AS-Acc | **80.63** | 80.34 | 80.15 |
| | AS-F1 | 73.04 | **73.22** | 72.84 |
| | I-F1 | **69.25** | 67.12 | 66.98 |
| Restaurant15 | AE-F1 | 70.15 | **70.26** | 69.87 |
| | AS-Acc | **83.66** | 82.41 | 82.11 |
| | AS-F1 | **58.57** | 58.39 | 58.02 |
| | I-F1 | **58.74** | 57.18 | 57.23 |

Note:
1. The test results are obtained under the condition that $\eta = 6$ and $\xi = 10$;
2. The results on the Laptop14 dataset is obtained when $\alpha = \beta = \gamma = 0.3$; the test results on the Restaurant14 and Restaurant15 datasets are obtained when $\alpha = 0.3, \beta = 0.4$ and $\gamma = 0.5$;
3. Bolded numbers are the highest values of the corresponding indicators.

4.5. **Case analysis.** Table 5 shows the opinion extraction results of some typical samples obtained by the two pipeline models, two ablation models and our proposed model. As the table shows, on Sample 1, Pipeline1 and the OURS-w/o-NNR model showed inconsistent sentiment tags for multiple aspect terms, while other models classified the aspect-level sentiments correctly and showed good consistency in sentiment judgment, indicating that all models performed well in detecting different aspects in the text, but the Pipeline1 and the joint model devoid of inter-task correlations underperformed in identifying the sentiments for multiple aspects. On Sample 2, Pipeline2 and OURS-w/o-FBR misclassified the aspect-level sentiment, which is because these two models assigned inconsistent sentiment tags to different aspect terms. On Sample 3, all other models except our model output the sentiment classification result as “neutral”, and on Sample 4, most models correctly classified the aspect-level sentiment as neutral. A further comparison of Sample 3 and Sample 4 shows that, though both samples are statements, Sample 3 implies positive sentiment and is likely to mislead the classifiers. It shows that all the models except ours have weak performance in identifying implicit sentiments in texts.

5. **Conclusion.** In this study, we constructed a practice-oriented joint model for aspect-level sentiment analysis. Specifically, based on the model we have built before [4], we migrated the aspect-level sentiment classification model onto the encoder-decoder architecture, and converted the model into a current word-based sequence sentiment tagging model, and thereby built a multi-task sequence tagging aspect-based sentiment analysis joint model. Meanwhile, as both the aspect extraction and sentiment classification tasks adopt the word-by-word decoding mode, we used the correlations between the aspect and sentiment of the current word as well as the correlations between the current word and the preceding word to improve the performance of the joint model. The experiment results show that our joint model outperformed other models in extracting the overall aspect-level sentiment features, which proved the validity of our joint model. Compared with the two pipeline models, our model performed better on the Restaurant dataset than on the

Table 5. Examples of sentiment extraction.

| Sample No. | PIPLINE1 | PIPLINE2 | OURS-w/o-NNR | OURS-w/o-FBR | OURS |
|---|---|---|---|---|---|
| 1 | food[pos]<br>service[neg] | food[pos]<br>service[pos] | food[neg]<br>service[pos] | food[pos]<br>service[pos] | food[pos]<br>service[pos] |
| 2 | Indian food[neg] | Indian food[pos] | Indian food[neg] | Indian food[pos] | Indian food[neg] |
| 3 | steak[neu] | steak[neu] | steak[neu] | steak[neu] | steak[pos] |
| 4 | decor[pos] | decor[neu] | restaurant[neu] | decor[neu] | decor[neu] |

1.Even when the chef is not in the house, the **food[pos]** and **service[pos]** are right on target
2.I know real Indian **food[neg]** and this was n't it
3.The **steak[pos]** melted in my mouth
4.The **decor[neu]** is designed in a contemporary Japanese style restaurant

Laptop dataset, indicating that the correlations in the joint model played a more salient role on the Restaurant dataset. It also reveals that our model needs to be improved to address classification tasks on comments of complex sentimental orientations, and we will try to make efforts in this regard in our future work.

## REFERENCES


[1] CHEN J, ZHANG J, XUE X, et al. Multi-view Ontology Alignment Visualization (MOAV): A Human-Cognition Based Ontology Alignment Visualization Technique [J]. *Journal of Network Intelligence*, 5(4): 198-210, 2020.

[2] LI X, BING L, LI P, et al. A Unified Model for Opinion Target Extraction and Target Sentiment Prediction[C],*Proceedings of the AAAI Conference on Artificial Intelligence.* 33: 6714–6721,2019.

[3] HE R, LEE W S, NG H T, et al. An Interactive Multi-Task Learning Network for End-to-End Aspect-Based Sentiment Analysis[C],*Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*,pp 504–515,2019.

[4] FAN S. X. et al. An Aspect Term Extraction Method Based on the Sequence Tagging Feedback Model [J]. *Computer Engineering and Design*, 41(09): 2643–2649, 2020.

[5] MA D, LI S, WU F, et al. Exploring Sequence-to-Sequence Learning in Aspect Term Extraction[C],*Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy: Association for Computational Linguistics*: 3538–3547 ,2019.

[6] HE R, LEE W S, NG H T, et al. An unsupervised neural attention model for aspect extraction[C], *ACL 2017 - 55th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers).* 1: 388–397,2017.

[7] XU H, LIU B, SHU L, et al. Double embeddings and cnn-based sequence labeling for aspect extraction[C], *ACL 2018 - 56th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers)*, 2: 592–598, 2018.

[8] YIN Y, WEI F, DONG L, et al. Unsupervised word and dependency path embeddings for aspect term extraction[C],*IJCAI International Joint Conference on Artificial Intelligence.* 2016-Janua: 2979–2985,2016.

[9] WANG W, PAN S J, DAHLMEIER D, et al. Coupled multi-layer attentions for co-extraction of aspect and opinion terms[C],*31st AAAI Conference on Artificial Intelligence*, AAAI 2017: 3316–3322, 2017.

[10] MA D, LI S, ZHANG X, et al. Interactive attention networks for aspect-level sentiment classification[C],*IJCAI International Joint Conference on Artificial Intelligence.*0: 4068–4074,2017.

[11] WANG B, LU W. Learning latent opinions for aspect-level sentiment classification[C],*32nd AAAI Conference on Artificial Intelligence,* AAAI 2018 : 5537–5544, 2018.

[12] TANG J, LU Z, SU J, et al. Progressive self-supervised attention learning for aspect-level sentiment analysis[C] *ACL 2019 - 57th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference.* : 557–566,2020.

[13] CHEN Z, QIAN T. Transfer capsule network for aspect level sentiment classification[C] *ACL 2019 - 57th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*: 547–556 ,2020.

[14] HE R, LEE W S, NG H T, et al. Exploiting document knowledge for aspect-level sentiment classification[C], *ACL 2018 - 56th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers). Melbourne, Australia: Association for Computational Linguistics*, 2: 579–585 ,2018.

[15] LUO H, LI T, LIU B, et al. DOER: Dual Cross-Shared RNN for Aspect Term-Polarity Co-Extraction[J]. *ArXiv*, abs/1906.0: 591–601 ,2019.

[16] ZHANG M, ZHANG Y, VO D T. Neural networks for open domain targeted sentiment[C] *Conference Proceedings - EMNLP 2015: Conference on Empirical Methods in Natural Language Processing.* : 612–621, 2015(September).

[17] MITCHELL M, AGUILAR J, WILSON T, et al. Open domain targeted sentiment[C],*EMNLP 2013 - 2013 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*: 1643–1654, 2013(October).

[18] HU M, PENG Y, HUANG Z, et al. Open-Domain Targeted Sentiment Analysis via Span-Based Extraction and Classification[J]. *ArXiv*,abs/1906.0: 537–546, 2019.

[19] PONTIKI M, GALANIS D, PAVLOPOULOS J, et al. SemEval-2014 Task 4: Aspect Based Sentiment Analysis[C], *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014). Dublin, Ireland: Association for Computational Linguistics and Dublin City University*: 27–35, 2014.

[20] PONTIKI M, GALANIS D, PAPAGEORGIOU H, et al. SemEval-2015 Task 12: Aspect Based Sentiment Analysis[C],*Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015). Denver, Colorado*: 486–495, 2015.

[21] BIRD S. NLTK: The Natural Language Toolkit[C], *Proceedings of the COLING/ACL on Interactive presentation sessions*: 69–72, 2006.

[22] CORENLP. Stanford CoreNLP , *Natural language software[EB/OL]*. Stanford, 2019.

[23] KINGMA D P, BA J L. Adam: A method for stochastic optimization[C], *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*: arXiv:1412.6980, 2015.