

SpanMTL: A Span-based Multi-table Labeling for Aspect-oriented Fine-grained Opinion Extraction

Yan Xing (✉ yxing425@163.com)

Civil Aviation University of China <https://orcid.org/0000-0001-5537-0883>

Yuexuan Zhu

Civil Aviation University of China

Wei Fan

Civil Aviation University of China

Yuxiang Zhang

Civil Aviation University of China

Rui Huang

Civil Aviation University of China

Zhaojun Gu

Civil Aviation University of China

W.H. Ip

The Hong Kong Polytechnic University

Kai-Leung Yung

The Hong Kong Polytechnic University

Research Article

Keywords: AFOE, span-level information, multi-table labeling, sentiment analysis

Posted Date: June 7th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1275207/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

SpanMTL: A Span-based Multi-table Labeling for Aspect-oriented Fine-grained Opinion Extraction

Yan Xing¹, Yuexuan Zhu^{1†}, Wei Fan^{1†}, Yuxiang Zhang^{1†}, Rui Huang^{1*†}, Zhaojun Gu^{1†}, W.H. Ip^{2†} and Kai-Leung Yung^{2†}

¹Department of Computer Science and Technology, Civil Aviation University of China, Tianjin, 300300, China.

²Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hong Kong, 999077, China.

*Corresponding author(s). E-mail(s): rhuang@cauc.edu.cn;
 Contributing authors: yxing@cauc.edu.cn; robbin0213@163.com;
wfan@cauc.edu.cn; yxzhang@cauc.edu.cn; zjgu@cauc.edu.cn;
mfwhip@inet.polyu.edu.hk; kl.yung@polyu.edu.hk;

[†]These authors contributed equally to this work.

Abstract

Aspect-oriented Fine-grained Opinion Extraction (AFOE) aims to extract the aspect terms, corresponding opinion terms and sentiment polarity in a target sentence. Most previous methods treat AFOE as word-level or span-level task, which ignore the complementarity of these two tasks. To integrate the merits of word-level and span-level information, we construct an end-to-end Span-based Multi-Table Labeling (SpanMTL) framework. SpanMTL combines word-based and span-based table labeling to tackle AFOE task. Specifically, in the proposed model, we use two separate BiLSTMs to encode the information of aspect and opinion terms into a word-based 2D representation table. Based on the table, we construct span-based table with CNN by associating the word-pair representations. At last, we integrate the table label distributions of word- and span-based table labeling to generate a multi-table labeling. The proposed method improves the performances of OPE and OTE tasks by introducing span information especially on the data with lots of spans. We have conducted various experiments on AFOE datasets to validate our

method. The experimental results show that our method outperforms other baselines when the sentences having lots of span information.

Keywords: AFOE, span-level information, multi-table labeling, sentiment analysis

1 Introduction

Aspect-based sentiment analysis (ABSA) [1–4] is a fine-grained task in sentiment analysis [5–7] that mainly focuses on aspect terms, opinion terms and sentiment polarity to obtain word-level or span-level sentiment information in sentences. Aspect terms are words or spans that describe the entities in a sentence, such as “waiters” and “pasta” in the sentence in Fig. 1. Opinion terms are words or spans that can reflect the subjective attitudes of the related aspect terms, such as “friendly” and “not good” in the sentence in Fig. 1. Sentiment polarity is the emotion attribute classification (positive, neutral, negative) based on aspect terms and related opinion terms. For example, the sentiment polarity of the aspect term “waiters” in the example sentence is “positive”.

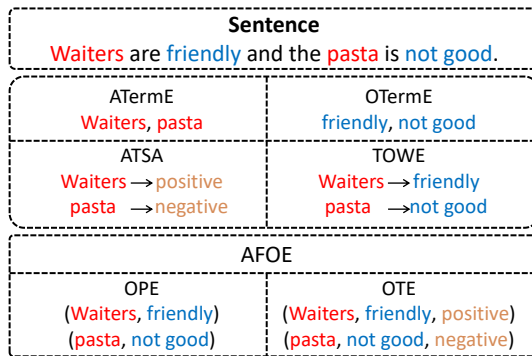


Fig. 1 An example of Aspect-oriented Fine-grained Opinion Extraction task.

As the ABSA task has received more and more attention from scholars, many sub-tasks around the above three elements have been proposed corresponding to the example shown in Fig. 1, which are explained as follows:

- Aspect Term Extraction (ATermE): extracting aspect term(s);
- Opinion Term Exaction (OTermE): extracting opinion term(s);
- Aspect Term Sentiment Analysis (ATSA): predicting the sentiment polarity of the annotated aspect term(s);

- Target-oriented Opinion Words Extraction (TOWE): extracting the opinion term(s) of annotated aspect term(s);
- Opinion Pair Extraction (OPE): extracting the aspect term(s) and corresponding opinion term(s);
- Opinion Triplet Extraction (OTE) or Aspect Sentiment Triplet Extraction (ASTE): extracting the aspect term(s) and related opinion term(s), and predicting the corresponding sentiment polarity;
- Aspect-oriented Fine-grained Opinion Extraction (AFOE): extracting aspect-opinion pair(s) or aspect-opinion-sentiment triplet(s).

The basic subtasks of ABSA, such as ATermE and OTermE, cannot fully analyze the fine-grained sentiment information in sentences. Therefore, Peng et al. [8] and Zhao et al. [9] propose OTE task and OPE task respectively. And then, Wu et al. [10] combine these two tasks into an AFOE task and provide four AFOE datasets. Different from most of the previous methods, which are based on the perspective of co-extraction with joint models [8, 11], the pipeline methods are built to first extract aspect terms and opinion terms, then pair them. This pipeline framework is intuitive, but it ignores the relationship between aspect terms and opinion terms, which might result in error propagation problem.

To overcome such drawbacks, many end-to-end approaches are proposed to solve OPE [9] and OTE [12–16] tasks. However, most of the previous methods only focus on solving OPE or OTE task separately. Although Wu et al. [10] present a novel labeling method, GTS, to solve OPE and OTE tasks simultaneously, GTS is not good at extracting pairs or triplets containing spans. SpanMlt [9] pays attention to span extraction in solving OPE task. However, most existing methods define OPE and OTE as word-level tasks, ignoring the information of spans. The interaction between spans is ignored in the span encoding stage. In addition, words and spans are not distinguished in the final extraction stage. These result in suboptimal results of AFOE when many aspect terms and opinion terms in a sentence are spans.

In this paper, we propose a Span-based Multi-Table Labeling method, called SpanMTL, which aims to extract aspect-opinion pair(s) or aspect-opinion-sentiment triplet(s) in a sentence through a 2D table labeling scheme on word and span tables. For word encoding, we use two separate BiLSTMs to simultaneously obtain aspect terms information and opinion terms information. Then we build a word-based 2D representation table, which is composed of aspect and opinion word representations. For table encoding, we use multi-dimensional recurrent neural network (MDRNN) [17] to learn the interactive information between elements in 2D table to obtain a contextual pair representation. For span encoding, we use the pair representation learned from the 2D word table to obtain the span pair representation through a convolutional neural network (CNN), and obtain a span-based 2D table. For decoding, we use table labeling method [10] to label the word- and span-based 2D tables. Finally, we integrate the labeling results of both tables and take the average value as

the final result. We have conducted various experiments on four benchmark datasets to compare our SpanMTL model with other baseline models. The experimental results show that our proposed model outperforms other state-of-the-art methods on the 15res and 16res datasets, which have lots spans. The main contributions of our method are summarized as follows:

- We generate word representation learned from two separate BiLSTMs, which can encode words into aspect- and opinion-specific word representations.
- We design a CNN based span encoder to convert the word pair representation into span pair representation to extract span information.
- Our method can employ both the word- and span-level information to improve the performance of AFOE task.

2 Related work

Aspect-based sentiment analysis (ABSA) was first proposed by Hu and Liu [1] and defined as an aspect-level task, which aims to mine the emotion information of the aspect term in a sentence. ABSA is also called fine-grained opinion mining, which mainly includes aspect term extraction (ATermE) [18–20], opinion term extraction (OTermE) [21], aspect term sentiment analysis (ATSA) [22–25], target-oriented opinion words extraction (TOWE) [26], etc. These subtasks dig out different information, but they do not integrate these useful information into one task.

In order to extract comprehensive aspect-level information from unmarked data, opinion pair extraction (OPE) task and opinion triplet extraction (OTE) task are proposed. For OPE task, Zhao et al. [9] propose a span-based multi-task end-to-end model, which first obtains representations of all words and spans and then predicts their labels and relationships. Chen et al. [16] and Mao et al. [15] transform OPE task into a machine reading comprehension task. Zhang et al. [12] tackle OPE task by a multi-task learning framework. For OTE task, Peng et al. [8] first propose a two-step method. The first step is to extract aspect terms and sentiment polarity through a unified labeling model, and then to extract opinion terms through a graph convolutional neural network. The second step is to use a binary classifier to determine all candidate triplets. From the perspective of first extracting and then judging, Huang et al. [11] also propose a novel two-step method. However, this two-step framework suffers from the problem of error propagation. To overcome this flaw, most recent works employ an end-to-end framework to solve the OTE task. Xu et al. [14] design a novel position-aware tagging scheme that is capable of jointly extracting the triplets.

In order to comprehensively evaluate OPE task and OTE task, Wu et al. [10] collectively refer to these two tasks as the aspect-oriented fine-grained opinion extraction (AFOE) and propose a novel tagging scheme, Grid Tagging Scheme (GTS). This method can mark the relationship between all word pairs in a word-based table. However, it cannot effectively extract aspect terms or opinion terms when these terms are spans. In addition, using one word

encoder to obtain both aspect and opinion word representation means that the difference between aspect and opinion words is not considered. In contrast, we propose a Span-based Multi-Table Labeling method (SpanMTL), which can explore both information of word- and span-level information to improve the performance of AFOE task.

3 Methodology

3.1 Problem Definition

Given a sentence $C = (w_1, w_2, \dots, w_c)$, where c is the number of words. The goal of opinion pair extraction (OPE) is to extract a set of aspect-opinion (A-O) pair(s) in C :

$$P = \{(a^j, o^j)\}_{j=1}^P, \quad (1)$$

where (a^j, o^j) is the j -th aspect-opinion pair, a is the aspect term, o is the corresponding opinion term, and P is the number of opinion pairs. The goal of opinion triplet extraction (OTE) is to extract a set of aspect-opinion-sentiment (A-O-S) triplet(s) in C :

$$T = \{(a^j, o^j, s^j)\}_{j=1}^T, \quad (2)$$

where (a^j, o^j, s^j) is the j -th aspect-opinion-sentiment triplet, a is the aspect term, o is the corresponding opinion term, s is the sentiment polarity and $s \in \{Positive, Neutral, Negative\}$, T is the number of triplets.

Note that there are more than one aspect term and opinion term in a sentence, an aspect term can be related to more than one opinion term, and the same opinion term may also correspond to multiple aspect terms, resulting in more than one A-O pair and A-O-S triplet in a sentence.

3.2 Table Labeling Scheme

We address the task of AFOE based on GTS [10], which utilizes two sets of target labels $\mathcal{Y}^P = \{A, O, P, N\}$ and $\mathcal{Y}^T = \{A, O, Pos, Neu, Neg, N\}$ to represent the relation of any word-pair (w_i, w_j) or span pair (s_m, s_n) in a sentence for the OPE and OTE task, respectively. Label “A” and “O” represent Aspect and Opinion respectively, which mean that the current word-pair or span-pair is an aspect term or opinion term. Label “P” indicates that the word-pair or span-pair is an A-O pair for OPE task. Label “Pos”, “Neu” and “Neg” represent that the word-pair or span-pair is an A-O-S triplet with a specific sentiment polarity for OTE task. While label “N” represents Other, indicating that the current word-pair or span-pair is irrelevant. Labeling examples of OPE and OTE task is shown in Fig. 2. Note that in order to make the extracted information more obvious, label “N” is not marked in Fig. 2.

After obtaining the final labels in the table, we extract the A-O pair(s) or A-O-S triplet(s). First, if the pair formed by a word itself or all pairs formed

by each word in a span in the table are marked as label “A” or “O”, then this word or span is extracted as an aspect term or opinion term. Second, after obtaining all the aspect terms and opinion terms, if a word-pair composed of aspect term and opinion term is marked as “P” or any sentiment polarity label (“Pos”, “Neu”, “Neg”), then this aspect term and opinion term are combined into an A-O pair or A-O-S triplet.

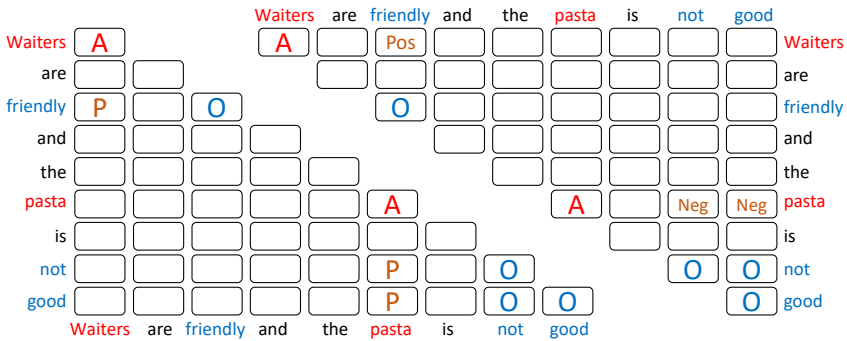


Fig. 2 Labeling examples for the OPE and OTE task. The lower triangular grid is OPE task and the upper triangular grid is OTE task.

3.3 Model Description

In this work, we propose an Span-based Multiple Table Labeling (SpanMTL) model to solve aspect-oriented fine-grained opinion extraction task. As shown in Fig. 3, our method mainly consists of three components: (1) **word encoding and word-based table encoding**, in which word pairs are composed of the embeddings of aspect words and opinion words learned by two independent word encoders, and then the influence of pairs and adjacent elements are captured by using a table encoder; (2) **span encoding and span-based table encoding**, in which related word pairs in the word-based table are used to form a span-based pair representation through a span encoder, and then the interacted span table is obtained through the table encoder; (3) **multiple table labeling**, including word table labeling, span table labeling, and the A-O pair or A-O-S triplet prediction.

In the A-O pair or A-O-S triplet prediction, we label the word-based table and the span-based table. Finally, we take the average probability of the results of the two tables as the final labeling result.

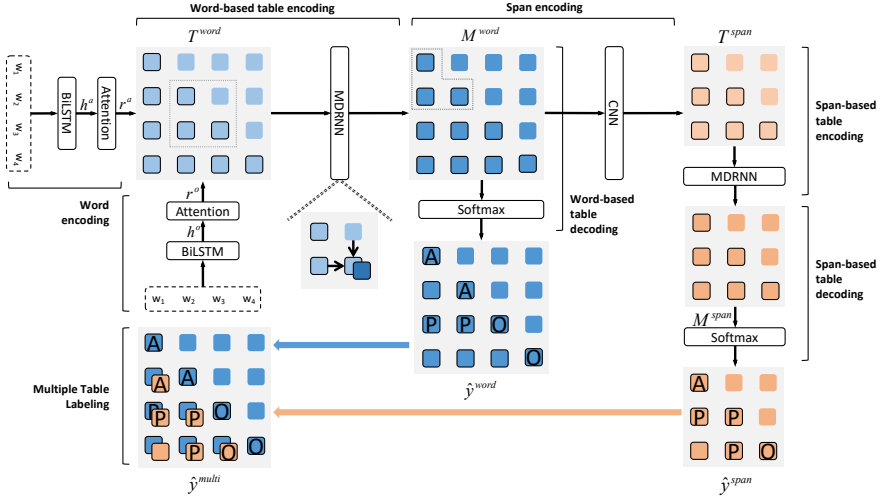


Fig. 3 An overview of SpanMTL for Aspect-oriented Fine-grained Opinion Extraction.

3.4 Word Encoding and Word-based Table Encoding

3.4.1 Word Encoding

In order to capture the semantic information of aspect words and opinion words in a sentence respectively, we design an aspect-specific word encoder and an opinion-specific word encoder. Each encoder is composed of a bidirectional Long Short-Term Memory (BiLSTM) layer and an attention layer. The inputs are a sentence with multiple words: $C = (w_1, w_2, \dots, w_c)$, where c is the number of words. We first encode all words in the sentence through BiLSTM layers separately to obtain the aspect-specific word representations h^a and opinion-specific word representations h^o with sequence features:

$$h_i^a = \left[\vec{h}_i^a, \overleftarrow{h}_i^a \right], \quad (3)$$

$$h_i^o = \left[\vec{h}_i^o, \overleftarrow{h}_i^o \right], \quad (4)$$

where \vec{h}_i is the forward hidden state and \overleftarrow{h}_i is the backward hidden state. Then we use the attention mechanism to focus on obtaining the aspect and opinion words information in the sentence and get the word representation r^a and r^o .

3.4.2 Word-based Table Encoding

Different with GTS [10] that obtains the word pair representation from a single encoder, we use the word representation learned by two independent encoders to form the word pair. Then, we build a 2D word-based table T^{word} , which is a matrix with dimension $c * c$ and contains all possible word pairs. The pair

representation of the i -th word as the aspect word and the j -th word as the opinion word is defined as follows:

$$\mathbf{T}_{i,j}^{word} = \mathbf{r}_i^a \oplus \mathbf{r}_j^o. \quad (5)$$

After that, we design a table encoder to contextualize the current element through surrounding elements in two directions, i.e. upper and left. The table encoder uses a Multi-Dimensional Recurrent Neural Network (MDRNN) [17] with Gated Recurrent Unit (GRU) [27] to enrich the 2D table elements representation $\mathbf{T}_{i,j}^{word}$, and obtain the interacted table \mathbf{M}^{word} . Thus, the new representation of word pairs is calculated by:

$$\mathbf{M}_{i,j}^{word} = GRU \left(\mathbf{T}_{i,j}^{word}, \mathbf{T}_{i-1,j}^{word}, \mathbf{T}_{i,j-1}^{word} \right), \quad (6)$$

where $\mathbf{T}_{i-1,j}^{word}$ and $\mathbf{T}_{i,j-1}^{word}$ are the upper and left element of the current pair $\mathbf{T}_{i,j}^{word}$ in 2D table, respectively.

3.5 Span Encoding and Span-based Table Encoding

3.5.1 Span Encoding

In order to solve the problem that existing methods cannot effectively obtain and extract span information, we design a span encoder composed of a convolutional neural network (CNN), which aims to convert the word pair representation in the word-based table into a span pair representation $\mathbf{T}_{m,n}^{span}$:

$$\mathbf{T}_{m,n}^{span} = conv \left(\mathbf{M}_{i-1,j}^{word}, \mathbf{M}_{i,j+1}^{word} \right), \quad (7)$$

where $\mathbf{M}_{i-1,j}^{word}$ and $\mathbf{M}_{i,j+1}^{word}$ are the upper and right word pairs of the current position. As shown in Fig. 4, this setting allows more comprehensive word-level information for span pairs.

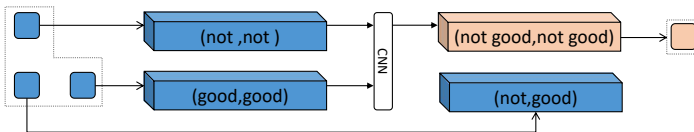


Fig. 4 A case of span encoder.

3.5.2 Span-based Table Encoding

After obtaining the representation of span pairs, the table encoder with MDRNN is used to model the interactions between different span pairs and

obtain the interacted table \mathbf{M}^{span} . The pair in m -th row and n -th column is defined as:

$$\mathbf{M}_{m,n}^{span} = GRU(\mathbf{T}_{m,n}^{span}, \mathbf{T}_{m-1,n}^{span}, \mathbf{T}_{m,n-1}^{span}), \quad (8)$$

where $\mathbf{T}_{m-1,n}^{span}$ and $\mathbf{T}_{m,n-1}^{span}$ are the upper and left elements of the current pair $\mathbf{T}_{m,n}^{span}$. The span-based table encoding layer is the same as that of word.

3.6 Multiple Table Labeling

After two table encoding layers based on word and span, we obtain the final representation $\mathbf{M}_{i,j}^{word}$ and $\mathbf{M}_{m,n}^{span}$ for each word pair and span pair. Then predict the table label distribution $\hat{y}_{i,j}^{word}$ and $\hat{y}_{m,n}^{span}$ as follows:

$$\hat{y}_{i,j}^{word} = \text{Softmax}\left(W^{word}\mathbf{M}_{i,j}^{word} + b^{word}\right), \quad (9)$$

$$\hat{y}_{m,n}^{span} = \text{Softmax}\left(W^{span}\mathbf{M}_{m,n}^{span} + b^{span}\right). \quad (10)$$

The cross entropy loss of A-O pair or A-O-S triplet prediction for word-based table and span-based table is:

$$\mathcal{L}^{word} = -\sum_{i=1}^c \sum_{j=i}^c \sum_{k \in \mathcal{Y}} \mathbb{I}(y_{i,j} = k) \log\left(\hat{y}_{i,j|k}^{word}\right), \quad (11)$$

$$\mathcal{L}^{span} = -\sum_{m=1}^{c-1} \sum_{n=m}^{c-1} \sum_{k \in \mathcal{Y}} \mathbb{I}(y_{m,n} = k) \log\left(\hat{y}_{m,n|k}^{span}\right), \quad (12)$$

where $y_{i,j}$ and $y_{m,n}$ are the ground truth distribution of A-O pair or A-O-S triplet, $\mathbb{I}(\cdot)$ is the indicator function. \mathcal{Y} is the label set (\mathcal{Y}^P or \mathcal{Y}^T) which is mentioned in Subsection 3.2, \mathcal{Y}^P for OPE task and \mathcal{Y}^T for OTE task.

To integrate the labels of the word-based table and span-based table, as shown in the matrix in the lower left corner of Fig. 3, we adopt the strategy of average probability to obtain the final results $\hat{y}_{i,j}^{multi}$:

$$\hat{y}_{i,j}^{multi} = (\hat{y}_{i,j}^{word} + \hat{y}_{m,n}^{span}) / 2, \quad (13)$$

where $i = m + 1, j = n$ and $m, n \in (1, c - 1), i, j \in (1, c)$.

The final loss of our model for a sentence C is a weighted sum of \mathcal{L}^{word} and \mathcal{L}^{span} with L2-regularization term as follows:

$$\mathcal{L} = \lambda_1 \mathcal{L}^{word} + \lambda_2 \mathcal{L}^{span} + \lambda_3 \|\theta\|^2. \quad (14)$$

4 Experiments

4.1 Datasets and Experimental Setting

We utilize four AFOE datasets built by Wu et al. [10] which are aligned by TOWE datasets [26] and SemEval Challenge datasets [2–4] to comprehensively

evaluate the effectiveness of our proposed model. The summary statistics for the AFOE datasets are shown in Table 1. Precision, recall and F1 score are used as the metrics for evaluation.

For the word encoder, we use pre-trained word vectors with double embeddings of DE-CNN [20] which is composed of 300-dimensional GloVe [28] embeddings and 100-dimension fastText [29] embeddings. The dimension of LSTM cell is set to 50 and the kernel size of CNN on span encoder embedding is 2. For the table encoder, we use MDRNN with GRU. For training, we use Adam [30] to optimize and set the learning rate to 0.005, batch size to 32 and dropout rate to 0.5. Our model is conducted on four testing sets and the average result of five runs is reported.

Table 1 Statistics of AFOE datasets.

Datasets		Aspect		Opinion	
		Number	Percentage(%)	Number	Percentage(%)
14res	word	618	72.62	778	90.78
	Span	233	27.38	79	9.22
	Total	851	100.00	857	100.00
14lap	word	277	59.31	408	86.26
	Span	190	40.69	65	13.74
	Total	467	100.00	473	100.00
15res	word	298	69.95	405	87.28
	Span	128	30.05	59	12.72
	Total	426	100.00	464	100.00
16res	word	317	71.72	431	89.79
	Span	125	28.28	49	10.21
	Total	442	100.00	480	100.00

4.2 Comparative Methods

Since there are less existing works on AFOE, we compare our model with 12 combined models to evaluate the performance. Note that some of the comparative methods are designed only for OPE task or OTE task.

- **CMLA+Dis-BiLSTM.** CMLA [31] is an end-to-end model with a coupled multi-layer attention network. Dis-BiLSTM is built by BiLSTM network for relation extraction. CMLA+Dis-BiLSTM uses CMLA to co-extract aspect terms and opinion terms, and uses Dis-BiLSTM to pair them.
- **CMLA+C-GCN.** C-GCN [32] is an extension of graph convolutional network (GCN) which is tailored for relation extraction. CMLA+C-GCN uses

CMLA [31] to co-extract aspect terms and opinion terms, and pairs them by C-GCN.

- **RINANTE+C-GCN**. This model uses RINANTE [33] model to co-extract aspect terms and opinion terms, and then pairs them by C-GCN [32]. Specifically, RINANTE is an algorithm to automatically mine extraction rules from existing training examples based on dependency parsing results.
- **BiLSTM-ATT+Distance**. The model is a pipeline neural network that uses BiLSTM and attention mechanism for Aspect Term Extraction (ATermE) and uses the Distance [1] approach for Target-oriented Opinion Words Extraction (TOWE).
- **BiLSTM-ATT+Dependency**. Different from BiLSTM-ATT+Distance model, this model uses the Dependency [34] model for TOWE.
- **BiLSTM-ATT+IOG**. The model firstly establishes a BiLSTM-ATT network for ATermE, then uses the IOG [26] model with multi-direction LSTM to detect pairs or triplets.
- **DE-CNN+IOG**. This model is composed of DE-CNN [20] for ATermE and the IOG [26] approach for TOWE.
- **RINANTE+IOG**. This model is composed of RINANTE [33] for ATermE and the IOG [26] approach for TOWE.
- **GTS-BiLSTM** [10]. The model is an end-to-end table labeling approach based on Grid Tagging Scheme (GTS). Specifically, it firstly uses BiLSTM word encoder to generate the representation of each word, and then designs the inference strategy in GTS to exploit indications between all pairs.
- **GTS-CNN** [10]. Different from GTS-BiLSTM model, this model uses CNN in the encoding layer.
- **Li-unified-R+PD**. The model firstly extracts aspect terms and opinion terms, and predicts sentiment polarity by a unified tagging scheme proposed by [35]. Then it detects the relation by the Pair relation Detection (PD) proposed by [8].
- **Peng-unified-R+PD** [8]. The model is a two-step model, which firstly extracts aspect terms and predicts sentiment of aspect, and then encodes all candidate pairs for the final classification.
- **Peng-unified-R+IOG**. The model first extracts aspect terms and predicts sentiment of aspect, and then pairs them by a binary classification with softmax layer. Peng-unified-R+IOG uses the first step of Peng et al. [8] to extract the (aspect, sentiment) pair, then combines them with IOG [26].
- **IMN+IOG**. The model uses IMN [36] model to solve the task of aspect term and opinion term co-extraction and aspect-level sentiment classification, and then uses IOG [26] model to pair the extracted terms.

4.3 Result Analysis

We compare the results of our model with the baseline models on four datasets. Since the AFOE task is divided into two parts, i.e. OPE and OTE, the experimental results are shown in Table 2 and Table 3. The best and

second-best results are respectively in bold and underline. Gain reports the relative improvements between SpanMTL and the best/second-best performance results.

For the opinion pair extraction (OPE) task, the F1-score of SpanMTL model increases to 71.84%, 67.87% and 72.67% on datasets 14res, 15res and 16res compared with the second-best being 71.74%, 65.39% and 71.42%. This improvement indicates that our model focusing on span extraction and the impact information between pairs is very helpful to the OPE task. However, it is observed that the experimental result of our model on 14lap dataset is the second-best, lower than the best GTS-CNN model with 0.86% F1-score decline. This maybe that 14lap dataset belongs to the domain of lamps instead of the restaurant domain of the other three datasets.

Compared with the OPE task, the opinion triplet extraction (OTE) task not only needs to extract the A-O pair, but also needs to predict the sentiment polarity of the aspect term. For the OTE task, the F1-scores of our model increase by 0.81% and 2.33% on 15res and 16res datasets compared with the second-best values, while decrease by 0.45% and 0.48% on 14lap and 14res datasets compared with the best values. This slight decrease on 14res and 14res datasets could be caused by the unnecessary non-phrase information brought in the representation of all spans in the input sentence.

In summary, due to the learning of span information and the table encoding, our model can effectively extract A-O pairs and A-O-S triplets. However, our model has a little dependence on the context domain, resulting in lower results in the dataset 14lap, which is also the direction to be improved in our future work.

4.4 Results of ATermE and OTermE task

To further analyze the effects of different methods, we compare the performance of two subtasks, Aspect Term Extraction (ATermE) and Opinion Term Extraction (OTermE). In order to ensure the fairness of the results, we follow the best method [10] and compare the results on res14 and res15 datasets.

The results of ATermE and OTermE are shown in Table 4. On one hand, it can be observed that the F1 scores of SpanMTL on ATermE and OTermE are both improved on res15 dataset, indicating that SpanMTL is also helpful to the subtasks in solving AFOE task. On the other hand, the result of ATermE task is better than that of OTermE. This maybe because that the ratio of span-based aspect terms is more than the ratio of span-based opinion terms, and our method is designed to solve the problem of span labeling. This illustrates the effectiveness of SpanMTL in solving subtasks ATermE and OTermE.

4.5 Ablation Study

In order to further verify the performance of the table encoder and span processing method under our full model, we conduct an ablation study. All results are shown in Table 5 and Table 6, where “SpanMTL w/o table” is a variant

Table 2 Quantitative comparison of different methods on OPE task.

Dataset	14res			14lap			15res			16res		
	<i>P</i> (%)	<i>R</i> (%)	<i>F1</i> (%)	<i>P</i> (%)	<i>R</i> (%)	<i>F1</i> (%)	<i>P</i> (%)	<i>R</i> (%)	<i>F1</i> (%)	<i>P</i> (%)	<i>R</i> (%)	<i>F1</i> (%)
CMLA+Dis-BiLSTM	77.21	52.14	62.24	59.47	45.23	51.17	64.86	44.33	52.47	66.29	50.82	57.33
CMLA+C-GCN	72.22	56.35	63.17	60.69	47.25	53.03	64.31	49.41	55.76	66.61	59.23	62.70
RINANTE+C-GCN	71.07	59.45	64.69	<u>67.38</u>	52.10	58.76	65.52	42.74	51.73	-	-	-
BiLSTM-ATT+Distance	47.09	39.40	42.90	38.85	29.20	33.34	39.63	33.95	36.57	43.60	39.65	41.53
BiLSTM-ATT+Dependency	56.31	48.93	52.36	31.58	28.84	30.15	58.26	42.19	48.94	64.48	48.85	55.59
BiLSTM-ATT+IOG	69.99	61.58	65.46	64.93	44.56	52.84	59.14	56.38	57.73	66.07	62.55	64.13
DE-CNN+IOG	67.70	69.41	68.55	59.59	51.68	55.35	56.18	60.08	58.04	62.97	66.22	64.55
RINANTE+IOG	70.16	65.47	67.74	61.76	53.11	57.10	63.24	55.57	59.16	-	-	-
GTS-CNN	<u>74.13</u>	<u>69.49</u>	<u>71.74</u>	68.33	<u>55.04</u>	60.97	66.81	61.34	63.96	<u>70.48</u>	<u>72.39</u>	<u>71.42</u>
GTS-BiLSTM	71.32	67.07	69.13	61.53	54.31	57.69	<u>67.76</u>	<u>63.19</u>	<u>65.39</u>	70.32	70.46	70.39
SpanMTL	73.24	70.50	71.84	61.37	58.89	<u>60.11</u>	69.75	66.08	67.87	72.81	72.52	72.67
Gain	-3.97	1.01	0.10	-6.96	3.85	-0.86	1.99	2.89	2.48	2.33	0.13	1.25

Table 3 Quantitative comparison of different methods on OTE task.

Dataset	14res			14lap			15res			16res		
Method	<i>P</i> (%)	<i>R</i> (%)	<i>F1</i> (%)	<i>P</i> (%)	<i>R</i> (%)	<i>F1</i> (%)	<i>P</i> (%)	<i>R</i> (%)	<i>F1</i> (%)	<i>P</i> (%)	<i>R</i> (%)	<i>F1</i> (%)
Li-unified-R+PD	41.44	68.79	51.68	42.25	42.78	42.47	43.34	50.73	46.69	38.19	53.47	44.51
Peng-unified-R+P	44.18	<u>62.99</u>	51.89	40.40	<u>47.24</u>	43.50	40.97	54.68	46.79	46.76	62.97	53.62
Peng-unified-R+IO	58.89	60.41	59.64	48.62	45.52	47.02	51.70	46.04	48.71	59.25	58.09	58.67
IMN+IOG	59.57	63.88	61.65	49.21	46.23	47.68	55.24	52.33	53.75	-	-	-
GTS-CNN	70.79	61.71	65.94	55.93	47.52	51.38	60.09	<u>53.57</u>	<u>56.64</u>	62.63	66.98	64.73
GTS-BiLSTM	<u>67.28</u>	61.91	64.49	59.42	45.13	<u>51.30</u>	63.26	50.71	56.29	<u>66.07</u>	<u>65.05</u>	<u>65.56</u>
SpanMTL	66.91	64.14	<u>65.49</u>	<u>56.07</u>	46.61	50.90	<u>61.39</u>	53.99	57.45	72.17	64.09	67.89
Gain	-3.88	-4.65	-0.45	-3.35	-0.91	-0.48	-1.87	0.42	0.81	6.10	-2.89	2.33

Table 4 Quantitative comparison (F1-score) of different methods on ATermE and OTermE task. The abbreviations “A” and “O” respectively denote the ATermE and OTermE.

Methods	14res		15res	
	A	O	A	O
BiLSTM-ATT	79.03	80.55	73.59	73.01
DE-CNN	81.90	80.57	75.24	73.07
CMLA	81.22	80.48	76.03	74.67
RINANTE	81.34	83.33	73.38	75.40
GTS-CNN	81.82	83.07	77.33	75.23
GTS-BiLSTM	81.10	82.62	78.44	75.63
SpanMTL	82.62	82.30	79.76	75.93

Table 5 Ablation study on OPE task.

Dataset		Method		
		SpanMTL w/o table	SpanMTL w/o span	SpanMTL
14res	<i>P</i> (%)	69.89	73.38	73.24
	<i>R</i> (%)	69.19	61.82	70.50
	<i>F1</i> (%)	69.54	67.11	71.84
14lap	<i>P</i> (%)	66.07	60.78	61.37
	<i>R</i> (%)	52.37	51.19	58.89
	<i>F1</i> (%)	58.41	55.57	60.11
15res	<i>P</i> (%)	67.84	67.75	69.75
	<i>R</i> (%)	62.98	59.30	66.08
	<i>F1</i> (%)	65.32	63.25	67.87
16res	<i>P</i> (%)	71.10	70.27	72.81
	<i>R</i> (%)	71.24	68.91	72.52
	<i>F1</i> (%)	71.17	69.59	72.67

of our SpanMTL that removes the word-based table encoding and span-based table encoding, and “SpanMTL w/o span” is another variant of our SpanMTL that removes the span representation learning, the span-based table encoding and the span extraction.

Removing the table encoding process results in performance degradation, indicating that table encoding between aspect-opinion pairs or triplets is useful for prediction. Similarly, the F1-score of “SpanMTL w/o span” is also decreased, which further shows that our emphasis on span is correct, and the designed span-based representation learning and span-based table encoding can improve the AFOE task.

Table 6 Ablation study on OTE task.

Dataset		Method		
		SpanMTL w/o table	SpanMTL w/o span	SpanMTL
14res	<i>P</i> (%)	65.88	63.87	66.91
	<i>R</i> (%)	63.79	63.74	64.14
	<i>F1</i> (%)	64.82	63.80	65.49
14lap	<i>P</i> (%)	52.47	51.05	56.07
	<i>R</i> (%)	44.77	44.40	46.61
	<i>F1</i> (%)	48.32	47.49	50.90
15res	<i>P</i> (%)	64.17	59.95	61.39
	<i>R</i> (%)	48.74	49.28	53.99
	<i>F1</i> (%)	55.65	54.09	57.45
16res	<i>P</i> (%)	71.24	65.03	72.17
	<i>R</i> (%)	61.36	63.90	64.09
	<i>F1</i> (%)	65.91	64.46	67.89

4.6 Case Study

To further illustrate the performance of our module, we present a case study on OPE task comparing our model with exhaustive model GTS [10]. As shown in Table 7, Sentence #1 is selected to evaluate the effect of models on the span extraction, and Sentence #2 is selected to evaluate the effect of models on the span extraction and multiple A-O pairs extraction.

For Sentence #1, our model can correctly extract the aspect term and A-O pair, while GTS model only extracts the first word of the aspect term. This proves that our improved table labeling method can effectively extract phrases. For Sentence #2, our model successfully extracts the opinion word “great surprises”, but GTS model only extracts the word “great”. In our analysis, this result may be caused by the ignorance of span information in the GTS tagging method. It is worth noting that there are three A-O pairs in Sentence #2, but both of our model and GTS model only extract two pairs. This shows that our method has insufficient processing effect on multiple A-O pairs extraction, which will be the research direction of our future work.

5 Conclusions

In this paper, we propose a Span-based Multi-Table Labeling framework, SpanMTL, for aspect-oriented fined-grained opinion extraction (AFOE) task. Our method encodes all words in each sentence through two BiLSTMs to obtain aspect- and opinion-specific word representations. Then we obtain

Table 7 Two examples for the case study on OPE task.

#	Sentences	Golden	GTS	SpanMTL
1	Beef noodle soup is good as well.	(Beef noodle soup, good)	(Beef, good)	(Beef noodle soup, good)
2	The seafood is amazing, there's a good wine list, and the ever-changing menu always offers some great surprises.	(seafood, amazing), (wine list, good), (menu, great surprises)	(seafood, amazing), (menu, great)	(seafood, amazing), (menu, great surprises)

word-based table encoding with a multi-dimensional recurrent neural network (MDRNN). Based on the word-based table, we generate span encoding with CNN, which is further processed by another MDRNN to generate span-based table encoding. The final extractions are obtained by integrating the results of word-based table labeling and span-based table labeling. Experimental results show that our model outperforms the state-of-the-art methods on res15 and res16 datasets, which have lots of span information. In our future work, we will explore the relationship between multiple pairs or triplets in a sentence to improve the performance of AFOE task.

Acknowledgements

This work was partially supported by grants from the National Natural Science Foundation of China (No. U1933114, No. 62172418) and Open Project Foundation of Information Security Evaluation Center of Civil Aviation, China (No. ISECCA-202005).

References

- [1] Hu, M., Liu, B.: Mining and summarizing customer reviews. In: Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 168–177 (2004)
- [2] Pontiki, M., Galanis, D., Pavlopoulos, J., Papageorgiou, H., Androutsopoulos, I., Manandhar, S.: SemEval-2014 task 4: Aspect based sentiment analysis. In: Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), pp. 27–35. Association for Computational Linguistics, Dublin, Ireland (2014). <https://doi.org/10.3115/v1/S14-2004>. <https://aclanthology.org/S14-2004>
- [3] Pontiki, M., Galanis, D., Papageorgiou, H., Manandhar, S., Androutsopoulos, I.: Semeval-2015 task 12: Aspect based sentiment analysis. In: Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), pp. 486–495 (2015)

- [4] Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., Al-Smadi, M., Al-Ayyoub, M., Zhao, Y., Qin, B., De Clercq, O., *et al.*: Semeval-2016 task 5: Aspect based sentiment analysis. In: International Workshop on Semantic Evaluation, pp. 19–30 (2016)
- [5] Liu, G., Wu, J., Zhou, Z.-H.: Key instance detection in multi-instance learning. In: Asian Conference on Machine Learning, pp. 253–268 (2012). PMLR
- [6] Feldman, R.: Techniques and applications for sentiment analysis. *Communications of the ACM* **56**(4), 82–89 (2013)
- [7] Bakshi, R.K., Kaur, N., Kaur, R., Kaur, G.: Opinion mining and sentiment analysis. In: 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), pp. 452–455 (2016). IEEE
- [8] Peng, H., Xu, L., Bing, L., Huang, F., Lu, W., Si, L.: Knowing what, how and why: A near complete solution for aspect-based sentiment analysis. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, pp. 8600–8607 (2020)
- [9] Zhao, H., Huang, L., Zhang, R., Lu, Q., Xue, H.: Spanmlt: A span-based multi-task learning framework for pair-wise aspect and opinion terms extraction. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp. 3239–3248 (2020)
- [10] Wu, Z., Ying, C., Zhao, F., Fan, Z., Dai, X., Xia, R.: Grid tagging scheme for aspect-oriented fine-grained opinion extraction. *arXiv preprint arXiv:2010.04640* (2020)
- [11] Huang, L., Wang, P., Li, S., Liu, T., Zhang, X., Cheng, Z., Yin, D., Wang, H.: First target and opinion then polarity: Enhancing target-opinion correlation for aspect sentiment triplet extraction. *arXiv preprint arXiv:2102.08549* (2021)
- [12] Zhang, C., Li, Q., Song, D., Wang, B.: A multi-task learning framework for opinion triplet extraction. *arXiv preprint arXiv:2010.01512* (2020)
- [13] Li, Y., Wang, F., Zhang, W., Zhong, S.-h., Yin, C., He, Y.: A more fine-grained aspect-sentiment-opinion triplet extraction task. *arXiv preprint arXiv:2103.15255* (2021)
- [14] Xu, L., Li, H., Lu, W., Bing, L.: Position-aware tagging for aspect sentiment triplet extraction. *arXiv preprint arXiv:2010.02609* (2020)
- [15] Mao, Y., Shen, Y., Yu, C., Cai, L.: A joint training dual-mrc framework for aspect based sentiment analysis. *arXiv preprint arXiv:2101.00816* (2021)

- [16] Chen, S., Wang, Y., Liu, J., Wang, Y.: Bidirectional machine reading comprehension for aspect sentiment triplet extraction. arXiv preprint arXiv:2103.07665 (2021)
- [17] Graves, A., Fernández, S., Schmidhuber, J.: Multi-dimensional recurrent neural networks. In: International Conference on Artificial Neural Networks, pp. 549–558 (2007). Springer
- [18] Wang, W., Pan, S.J., Dahlmeier, D., Xiao, X.: Recursive neural conditional random fields for aspect-based sentiment analysis. arXiv preprint arXiv:1603.06679 (2016)
- [19] He, R., Lee, W.S., Ng, H.T., Dahlmeier, D.: An unsupervised neural attention model for aspect extraction. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 388–397 (2017)
- [20] Xu, H., Liu, B., Shu, L., Yu, P.S.: Double embeddings and cnn-based sequence labeling for aspect extraction. arXiv preprint arXiv:1805.04601 (2018)
- [21] Liu, P., Joty, S., Meng, H.: Fine-grained opinion mining with recurrent neural networks and word embeddings. In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pp. 1433–1443 (2015)
- [22] Dong, L., Wei, F., Tan, C., Tang, D., Zhou, M., Xu, K.: Adaptive recursive neural network for target-dependent twitter sentiment classification. In: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (volume 2: Short Papers), pp. 49–54 (2014)
- [23] Tang, D., Qin, B., Liu, T.: Aspect level sentiment classification with deep memory network. arXiv preprint arXiv:1605.08900 (2016)
- [24] Tay, Y., Tuan, L.A., Hui, S.C.: Learning to attend via word-aspect associative fusion for aspect-based sentiment analysis. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 32 (2018)
- [25] 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)
- [26] Fan, Z., Wu, Z., Dai, X., Huang, S., Chen, J.: Target-oriented opinion words extraction with target-fused neural sequence labeling. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pp. 2509–2518 (2019)

- [27] Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y.: Learning phrase representations using rnn encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078 (2014)
- [28] Pennington, J., Socher, R., Manning, C.D.: Glove: Global vectors for word representation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1532–1543 (2014)
- [29] Bojanowski, P., Grave, E., Joulin, A., Mikolov, T.: Enriching word vectors with subword information. Transactions of the Association for Computational Linguistics **5**, 135–146 (2017)
- [30] Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
- [31] Wang, W., Pan, S.J., Dahlmeier, D., Xiao, X.: Coupled multi-layer attentions for co-extraction of aspect and opinion terms. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 31 (2017)
- [32] Zhang, Y., Qi, P., Manning, C.D.: Graph convolution over pruned dependency trees improves relation extraction. arXiv preprint arXiv:1809.10185 (2018)
- [33] Dai, H., Song, Y.: Neural aspect and opinion term extraction with mined rules as weak supervision. arXiv preprint arXiv:1907.03750 (2019)
- [34] Zhuang, L., Jing, F., Zhu, X.-Y.: Movie review mining and summarization. In: Proceedings of the 15th ACM International Conference on Information and Knowledge Management, pp. 43–50 (2006)
- [35] Li, X., Bing, L., Li, P., Lam, W.: A unified model for opinion target extraction and target sentiment prediction. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, pp. 6714–6721 (2019)
- [36] He, R., Lee, W.S., Ng, H.T., Dahlmeier, D.: An interactive multi-task learning network for end-to-end aspect-based sentiment analysis. arXiv preprint arXiv:1906.06906 (2019)

Statements and Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Data availability statement All data generated or analysed during this study are included in this published article:

Wu, Z., Ying, C., Zhao, F., Fan, Z., Dai, X., Xia, R.: Grid tagging scheme for aspect-oriented fine-grained opinion extraction. arXiv preprint arXiv:2010.04640 (2020)