JASTE: An Efficient Multi-Task Model for Aspect Sentiment Triplet Extraction

Yanjing Wu, Xiangyu Liu, Kuilong Liu, Changyuan Yang and Chongjiang Li
Alibaba Group, HangZhou, Zhejiang Province

Abstract. Aspect sentiment triplet extraction (ASTE) is a relative difficult and novel research, which is a subtask of aspect-based sentiment analysis (ABSA). ASTE is a task that extracts triplets of aspects being discussed, relevant opinion entities and sentiment polarities from a given sentence. Existing approaches mainly deal with this problem by pipeline or simple multi-task structure, which do not take full advantage of the strong correlation among the three elements of the triplet. In this work, we adopt two special tagging schemes, AOBIO and Pair Tagging Scheme (PTS), and propose an efficient end-to-end multi-task model named Joint Aspect Sentiment Triplet Extraction (JASTE) to address ASTE task. JASTE is composed of three modules: aspect and opinion extraction module, relation module and sentiment module. Specially, the relation module is designed to capture the relationship between aspect and opinion properly. The three modules interact with each other by sharing the same embedding. Extensive experimental results on different benchmark datasets show that JASTE can significantly outperform state-of-the-art performances.

Keywords. Aspect Sentiment Triplet Extraction, Multi-task Structure, End-to-End

1. Introduction

Extracting aspect, opinion or sentiment from a sentence is a challenging and important work of sentiment analysis. There are many works about aspect, opinion and sentiment extraction or co-extraction in natural language processing. These tasks are collectively referred to aspect-based sentiment analysis (ABSA). ABSA includes but not limited to aspect term extraction (ATE) [1], opinion term extraction (OTE) [2], aspect term sentiment classification (ASC) [3], aspect opinion co-extraction (AOE) [4] and aspect sentiment triplet extraction (ASTE) [5,6,7,8]. Compared with other subtasks in ABSA, ASTE is a relative novel task, and basically untouched until 2020 [5]. It extracts more information at the same time which makes the model more interpretable. For example, two triplets of ('price', 'reasonable', 'positive') and ('customer service', 'poor', 'negative') will be extracted from the given sentence, "The price is reasonable although the customer service is poor.". Despite of remarkable performance the existing approaches has achieved, two issues need to be addressed.

Firstly, the existing tagging mechanisms are not fully suitable for ASTE task. On the one hand, some of them use BIO tagging scheme 1 to represent aspect, opinion, sentiment independently, which does not consider the relationship among different elements in a sentiment triplet [8,6,5]. On the other hand, some of them use the unified

---

tagging scheme to represent aspect tag, opinion tag and sentiment tag simultaneously [7], which does not take into account the sequence and dependence between aspect term and opinion term.

Secondly, the existing researches lack a module to predict the relationship between aspect and opinion easily. They use heuristic rules to form aspect and opinion pair based only on sentiment module. The sentiment module is designed to predict the sentiment polarity of aspect-opinion pair, which is not suitable to capture the relation between aspect and opinion. In order to circumvent these issues mentioned above, we propose a multi-task model named Joint Aspect Sentiment Triplet Extraction (JASTE). In this model, it uses AOBIO tagging scheme to tag aspect and opinion. Then the model will use Pair Tagging Scheme (PTS) to identify the relationship between aspect and opinion, and classify their sentiment polarity. Except these two special tagging schemes, JASTE learns context representation from an encoding layer firstly, and then uses three modules: aspect and opinion extraction module, relation module and sentiment module to extract aspect, opinion and sentiment simultaneously. In summary, we make following contributions in this work:

- Two tagging schemes, AOBIO and PTS, are used to tag label for different modules in our model. Different from the existing tagging mechanisms, our tagging schemes take full account of the dependence, sequence and relationship of various elements in a sentiment triplet.
- Compared with other models of ASTE, we innovatively design the relation module to model the relationship between aspect and opinion expressions.
- We propose an efficient end-to-end multi-task model, JASTE, to deal with ASTE task. Our model uses three modules to capture the relationship among aspect term, opinion term and sentiment polarity of a sentiment triplet properly. Experimental results on different benchmark datasets show that our model significantly outperforms state-of-the-art performances.

2. Tagging Scheme

In this section, we first give the problem formulation of Aspect Sentiment Triplet Extraction (ASTE). Then we introduce two special tagging schemes, AOBIO and PTS.

2.1. Problem Formulation

Given an input sequence $X = \{x_1,...,x_N\}$, our target is to output the sentiment triplets $T = \{(a,o,c)_i\}_{i=0}^{|T|}$, where x_i is a word, N represents the length of sequence, a is an aspect term, o is an opinion term, and c is a sentiment polarity. a and o can be a single word or a phrase, and c is one of {No-Dep, Neu, Pos, Neg}.

2.2. AOBIO

According to previous works [12,13,14], we know that aspect and opinion terms have strong co-occurrence. Aspect and opinion expressions are both texts, so their recognitions can be viewed as two sequence labeling processes. Based on above observations, we combine aspect recognition with opinion recognition into one sequence

label identification, and adopt AOBIO tagging scheme which is consisted of seven labels {BA, IA, BO, IO, O, S, E} to classify each word of the sequence.

2.3. PTS

Position Tagging Scheme (PTS) has two tagging mechanisms, one for the relation module and the other one for the sentiment module. For simplicity, we call PTS for relation module as RPTS, and PTS for sentiment module as SPTS. RPTS uses relation score to measure the degree of relationship of any word-pair (x_i, x_j) in relation module. The score of 1 represents that the word-pair is related, the score of 0 represents that the word-pair is not related. However, in sentiment module, SPTS uses four tags of {No-Dep, Neu, Pos, Neg}2 to represents the polarity of any word-pair (x_i, x_j). Here the word pair (x_i, x_j) is unordered, so the word pair (x_i, x_j) and (x_j, x_i) have the same tag.

3. Model

In order to deal with ASTE task, we propose an efficient multi-task model, JASTE. Figure 1 illustrates the architecture of our model. Firstly, an input sentence will be transformed by encoding layer to a sequence of vector representations. Secondly, aspect and opinion extraction module uses BERT-CRF [15] to extract aspect and opinion terms; relation module predict relationship between aspect and opinion through self-attention mechanism; sentiment module uses a classification layer to predict polarity. Finally, joint learning is used to define loss. During inference phase, we generate sentiment triplets of a sentence according to the predicted result of JASTE.

3.1. Encoding Layer

In our model, we first tokenize sentences from dataset into sequences by BertTokenizer3, then add [CLS] at the beginning of the sequence and [SEP] at the end of the sequence. After this preprocessing, we can obtain the input sequence $X = \{x_1,...,x_N\}$ from a given sentence.

   Because of the success of BERT [16] in ABSA task, our model also adopts BERT as encoding layer. BERT employs a multi-layer bidirectional Transformer to generate the contextual representations $H = \{h_1,...,h_N\}$ of the given sequence, the dimension of h in H is D_h. So, $h \in R^{1 \times D_h}$, $H \in R^{N \times D_h}$.
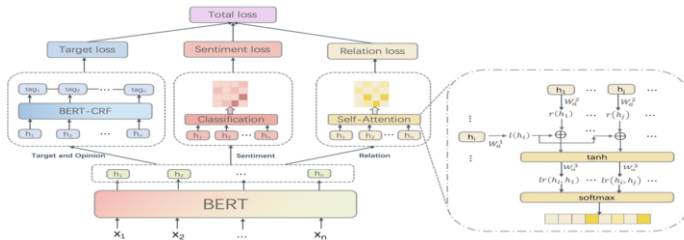

Figure 1. Architecture of Joint Aspect Sentiment Triplet Extraction

---

2 No-Dep, Neu, Pos and Neg are short names of no-dependence, neutral, positive and negative.
3 https://github.com/google-research/bert.git

3.2. Aspect and Opinion Extraction Module

Aspect and opinion extraction module is used to extract aspect and opinion terms from a given sequence. In order to deal with this sequence tagging task, we use BERT-CRF [15] to couple CRF upon the encoding layer. BERT-CRF adopts state probability matrix and transition probability matrix to model sequence labeling process. State probability matrix $P \in R^{N \times K}$ describes the relationship between words and labels which is calculated with a classifier. Transition probability matrix $Q \in R^{K \times K}$ indicates the relations between adjacent labels which is the parameter of CRF [17], where K represents the number of tags of AOBIO which is equal to 7. Overall, given an input sequence X, the probability of the predicted sequence $Y^{AO} = (y_i^{AO}, \dots, y_N^{AO})$ is calculated as follows:

$$p(Y^{AO}|X) = \frac{\exp(s(X, Y^{AO}))}{\Sigma_{\widehat{Y^{AO}} \in Y_X^{AO}} \exp(s(X, \widehat{Y^{AO}}))} \tag{1}$$

$$s(X, Y^{AO}) = \sum_{i=1}^{N} Q_{y_i^{AO}, y_{i+1}^{AO}} + \sum_{i=2}^{N} P_{i, y_i^{AO}} \tag{2}$$

$$P = HW_p + b_p \tag{3}$$

where y_X^{AO} represents all possible label sequences for the given sentence, $W_p \in R^{D_h \times K}$ and $b_p \in R^{N \times K}$ are the weight and bias of classifier in BERT-CRF. The hidden embedding H is the output of encoding layer. $P_{i, y_i^{AO}}$ represents the probability of the label of x_i is y_i^{AO}. $Q_{y_i^{AO}, y_{i+1}^{AO}}$ represents the probability of the label of x_{i+1} is y_{i+1}^{AO} after the label of x_i is y_i^{AO}.

3.3. Relation Module

In the existing researches of ASTE, they all extract aspect and opinion pairs based only on sentiment module. However, this method is not work in our tagging schemes. Because the model will get low recall if we extract aspect and opinion pairs based only on sentiment module. However, if we ignore these noises, the sentiment module will only focus on recognizing the sentiment polarity of word-pair. In order to solve this problem, our model designs the relation module to capture the relationship between aspect and opinion terms accurately. In section 4.5, the experimental result also prove that our relation module is an effective module.

As shown in Figure 1, the relation module uses the relation score to indicate the relationship between aspect and opinion, which makes it flexibly model token-level relations by adjusting the threshold. In order to compute the relation score of each wordpair, we design a self-attention mechanism. Suppose $r_{i,j}$ represents the relation score of a word-pair, then:

$$r_{i,j} = \frac{\exp(u_{i,j})}{\Sigma_{k=1}^{N} \exp(u_{i,k})} \tag{4}$$

$$u_{i,j} = tanh\left(W_{u3}\left(h_i W_{u1} + h_j W_{u2} + b_u\right)^T\right) \tag{5}$$

where $W_{u1}, W_{u2} \in R^{D_h \times D_u}$, $W_{u3} \in R^{1 \times D_u}$, $b_u \in R^{1 \times D_u}$. W_{u1}, W_{u2}, W_{u3} are weights of the self-attention layer, b_u is bias of the self-attention layer, D_u is the hidden dimension of the self-attention layer.

3.4. Sentiment Module

Unlike other existing researches of ASTE, our sentiment module only focus on predicting sentiment polarity of each word-pair. In order to solve this issue, we use a linear classification layer to classify labels of sentiment matrix. Assuming the gold sentiment matrix is Y^S, then the sentiment polarities probability distribution of each word-pair $p\left(Y_{i,j}^{S}\middle| x_i, x_j\right)$ can be calculated as follow:

$$p\left(Y_{i,j}^{S}\middle| x_i, x_j\right) = softmax\left(v_{i,j}\right) \tag{6}$$
$$v_{i,j} = e_{i,j}W_v + b_v \tag{7}$$
$$e_{i,j} = concat\left(\mathrm{h}_i; \mathrm{h}_j\right) \tag{8}$$

where $W_v \in \mathrm{R}^{(2\mathrm{D_h})\times \mathrm{D_v}}$, $b_v \in \mathrm{R}^{1\times \mathrm{D_v}}$. Wv and bv are weight and bias of the linear classification layer, Dv represents the number of labels in SPTS which is equal to 4. $p\left(Y_{i,j}^{S}\middle| x_i, x_j\right) \in \mathrm{R}^4$ represents probability distribution of different sentiment polarities of SPTS.

3.5. Joint Learning

JASTE is a model consisted of three modules: aspect and opinion extraction module, relation module and sentiment module. Each module is different, and represents one of subtasks of ASTE. It is unreasonable to simply sum up the loss of each module. Because different modules converge at different rates [18,19]. So we weighted sum the loss of each subtask to make them coverage at the same rate. The total loss of our model is formulated as follow:

$$L_{total} = \lambda_{AO}L_{AO} + \lambda_R L_R + \lambda_S L_S \tag{9}$$

$\lambda_{AO}, \lambda_R, \lambda_S$ are weights of subtask losses in JASTE. L_{AO}, L_R, L_S represents the losses of aspect and opinion module, relation module and sentiment module. For aspect and opinion module, we use the minimum likelihood loss function to get L_{AO}. Assuming that $\widetilde{Y^{AO}}$ is the gold label sequence for the given sentence X, then L_{AO} is calculated as:

$$L_{AO} = log \sum_{Y^{AO} \in Y_X^{AO}} ex\,p\left(s(X, Y^{AO})\right) - s\left(X, \widetilde{Y^{AO}}\right) \tag{10}$$

According to RPTS, there are only two different relation score, 0 or 1, in gold relation score matrix. So we use the cross entropy loss to get L_R:

$$L_S = -\sum_{i=1}^{N}\sum_{j=1}^{N}\left(\hat{p}\left(Y_{i,j}^{R}\middle| x_i, x_j\right)log\left[p\left(Y_{i,j}^{R}\middle| x_i, x_j\right)\right]\right) \tag{11}$$

Sentiment module is designed to classify the sentiment polarity of each word-pair properly. In order to avoid noises from abundant "No-Dep" tags, we use weighted crossentropy loss function to get L_S. We can avoid this noise by set the weight of "No-dep" to a small value such as 0.01. Then L_S is calculated as:

$$L_S = -\sum_{i=1}^{N}\sum_{j=1}^{N}\left(\hat{p}\left(Y_{i,j}^{S}\middle| x_i, x_j\right)log\left[p\left(Y_{i,j}^{S}\middle| x_i, x_j\right)\right]\right)W_S \tag{12}$$

Different from λ_S, $W_S \in \mathrm{R}^{4\times 1}$ represents the weights of different labels in sentiment module.

3.6. Inference

During inference phase, we extract three elements of a sentiment triplet simultaneously. In aspect and opinion extraction module, we use the Viterbi algorithm [17] to obtain an aspect set $A = \{a_1, a_2 ... a_{MA}\}$ and an opinion set $O = \{o_1, o_2 ... o_{MO}\}$, where M_A and M_O

represent the number of aspects and opinions. An aspect $a_i = \{x_{i_S^a},...,x_{i_E^a}\}$ in A or opinion $o_j = \{x_{j_S^o},...,x_{j_E^o}\}$ in O can be a single word or a phrase, where i_S^a and i_E^a represent the start and end positions of aspect a_i, j_S^o and j_E^o are the start and end positions of opinion o_j.

Whether an aspect a_i is related to an opinion o_j depends on their correlated degree $\alpha_{i,j}$. If $\alpha_{i,j}$ is higher than the given threshold $\tilde{\alpha}$, we think a_i and o_j are related and extract the pair (a_i, o_j). $r_{k,l}$ represents the relation score of a word-pair (x_k, x_l). Because word-pair is unordered, the correlated degree $\alpha_{i,j}$ between a_i and o_j can be calculated as:

$$\alpha_{i,j} = \frac{1}{2}\left(\frac{1}{i_E^a - i_S^a 2} \sum_{k=i_S^a}^{i_E^a} \sum_{l=j_S^o}^{j_E^o} r_{k,l} + \frac{1}{j_E^o - j_S^o} \sum_{l=j_S^o}^{j_E^o} \sum_{k=i_S^a}^{i_E^a} r_{l,k} \right) \tag{13}$$

If a_i and o_j has a relationship, the sentiment polarity of this pair is predicted by the probability distribution of sentiment module. We use $c_{i,j}$ to represent the sentiment polarity of (a_i, o_j), then a sentiment triplet $(a_i, o_j, c_{i,j})$ is generated. $c_{i,j}$ is calculated as follow:

$$c_{i,j} = argmax\left(\frac{1}{2}\left(\frac{1}{i_E^a - i_S^a} \sum_{k=i_S^a}^{i_E^a} \sum_{l=j_S^o}^{j_E^o} p\left(Y_{k,l}^S | x_k, x_l\right) + \right.\right.$$
$$\left.\left. \frac{1}{j_E^o - j_S^o} \sum_{l=j_S^o}^{j_E^o} \sum_{k=i_S^a}^{i_E^a} p\left(Y_{l,k}^S | x_l, x_k\right) \right) \right) \tag{14}$$

4. Experiments

4.1. Datasets

In order to prove the effect of our model, we conduct experiments on four benchmark datasets of semeval task: Rest14, Rest15, Rest16 and Lap14[4]. Rest14, Rest15 and Rest16 are datasets about restaurant domain, Lap14 is a dataset of laptop domain. Detailed statistics for these datasets are shown in Table 1.

4.2. Experimental Setting

In our model, we adopt the BERT-Base5 to output the bert embedding in encoding layer. The embedding dimension of self-attention layer is 100. To make input sentences have the same length, we set the sequence length to 64. We set the weight of losses λ_{AO}, λ_R, λ_S to 0.1, 0.5, 0.4. The weight of weighted cross-entropy loss [No-Dep, Neu, Pos, Neg] of sentiment module, WS, is [0.01, 1.0, 1.0, 1.0]. We use learing rate 2e-05 to train BERTBase and learning rate 0.001 to train other parameters in JASTE. We train our model for a maximal of 100 epochs with batch size 8 in GeForce GTX 1080 Ti. During inference phase, we set the relation threshold $\tilde{\alpha}$ to 0.5. Precision(P), recall(R), F1 score(F) are evaluation metrics in our work. It is worth mentioning that a gold triplet means that the aspect term, opinion term and sentiment polarity of it are all correct.

---

[4] https://github.com/databill86/SemEval-Triplet-data
[5] https://github.com/google-research/bert

Table 1. Statistics about the datasets. #S and #TR are the nums of sentences and sentiment triplets in datasets. #M shows the total num of triplets which contain multi-aspect triplets.

Dataset	Rest14			Lap14			Rest15			Rest16		
	#S	#TR	#M	#S	#TR	#M	#S	#TR	#M	#S	#TR	#M
train	1300	2145	2409	920	1265	1451	593	923	977	842	1289	1370
val	323	524	590	228	337	380	148	238	260	210	316	334
test	496	862	1014	339	490	552	318	455	479	320	465	507

Table 2. Results on different datasets (%). The highlighted scores in the table indicate the best results of test data.

Model	Rest14			Lap14			Rest15			Rest16		
	P	R	F	P	R	F	P	R	F	P	R	F
Pipeline	42.29	64.07	50.90	40.40	47.24	43.50	40.97	54.68	46.79	46.76	62.97	53.62
TASTE	54.7	54.7	54.7	38.6	38.6	38.6	54.7	54.7	54.7	54.3	54.3	54.3
OTE-MTL	66.04	56.25	60.62	50.52	39.71	44.31	57.51	43.96	49.76	64.68	54.97	59.36
JET	67.97	60.32	63.92	58.47	43.67	50.00	58.35	51.43	54.67	64.77	61.29	62.98
JASTE	72.17	68.68	70.38	60.60	55.06	57.70	62.50	63.41	62.95	68.17	68.31	68.24

4.3. Baselines

To better illustrate the effect of our model, we compare it with a variety of proposed models. On the one hand, we compare our model with baselines of pipeline structure. (1) Pipeline [5] captures triplets by GCN and then predicts a triplet is valid or not by multi-layer perceptron. (2) TASTE [6] extracts aspect term firstly, then identifies sentiment polarity relying on DTLSTM, and get the opinion span by HN or SS in the end.

On the other hand, we contrast JASTE with other multi-task models. (1) JETBERT [7] is a multi-task model with unified tag. It uses BERT as word embedding and extracts sentiment triplets simultaneously by position-aware tagging scheme. (2) OTEMTL [8] uses aspect tagging, opinion tagging and sentiment parsing these three modules to predict triplets.

4.4. Experimental Results

The predicted results of our model on different benchmark datasets are shown in Table 2. From the evaluated metrics, we can know that our model significantly outperforms other models in all metrics. The growth of F1 score illustrates that our model can capture the relationship among different elements in a sentiment triplet more properly, which mainly due to our model structure. The domains of laptop and restaurant are relative different. Existing models tend to get proper results just in the domain of restaurant, while our model can get decent evaluated metrics in both restaurant and laptop domains. This phenomenon states that our model has a better generalization.

4.5. Ablation Study

Most existing models, with the exception of JET, use GloVe [6] to initialize word embedding and adopt BiLSTM [21] as the encoding layer. In order to ensure the fairness of experiments, we add the experiment" +GloVe" to prove the effect of our model. Like other models, "+GloVe" uses GloVe to initialize the word embedding and adopts BiLSTM as the encoding layer. According to the results of Table 2 and Table 3, we know that "+GloVe" surpasses other models with GloVe and BiLSTM, which indicates that our model structure is effective. However, compared with original JASTE, F1 scores of "+GloVe" decrease, which means that BERT provides more information than GloVe and BiLSTM. JET [7] uses BERT to initialize the word embedding, but not adopts it as the encoding layer. To illustrate the effect of adopting BERT as the encoding layer, we add the experiment "−Tune". In "−Tune" model, we fix the parameters of BERT, regard the output of BERT as the word embedding and take BiLSTM as the encoding layer. According to Table 3, we know that using BERT instead of BiLSTM as the encoding layer can effectively improve the model metrics. Relation module is a very important module in our model, which is also a major difference between our model and other multi-task models. The model of "−Relation" removes the relation module, and predicts the relationship between aspect and opinion terms only depending on sentiment module. As shown in Table 3, the remarkable reduction of model metrics indicates that relation module can capture the relationship between aspect and opinion more properly than sentiment module.

Table 3. F1 scores of different models in ablation study.

Model	Rest14	Lap14	Rest15	Rest16
JASTE	70.38	57.70	62.95	68.24
+GloVe	62.59	46.12	55.78	60.62
−Tune	67.45	53.48	57.66	65.75
−Relation	53.27	50.21	51.47	51.79

5. Conclusion

In this work, we describe ASTE task in detail and analyze the existing approaches comprehensively. In order to deal with ASTE task properly, we propose a model named Joint Aspect Sentiment Triplet Extraction (JASTE). Our model adopts two special tagging schemes, AOBIO and PTS, to model the dependence and relationship among all elements in a sentiment triplet. JASTE is consisted of three modules of aspect and opinion module, relation module and sentiment module. Unlike other models, we specially design the relation module to capture the relationship between aspect and opinion. The experiments on all benchmark datasets of ASTE show that our model significantly outperforms state-of-the-art performances.

---

[6] https://nlp.stanford.edu/projects/glove

References

[1] Chernyshevich M. IHS R&D Belarus: Cross-domain extraction of product features using conditional random fields. 2014.

[2] Hu M, Liu B. Opinion extraction and summarization on the web. In: AAAI. vol. 7; 2006. p. 1621-4.

[3] Wang Y, Huang M, Zhu X, Zhao L. Attention-based LSTM for aspect-level sentiment classification. In: Proceedings of the 2016 conference on empirical methods in natural language processing; 2016. p. 606-15.

[4] Liu K, Xu L, Zhao J. Co-extracting opinion targets and opinion words from online reviews based on the word alignment model. IEEE Transactions on knowledge and data engineering. 2014;27(3):636-50.

[5] 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: AAAI; 2019. p. 8600-7.

[6] Sutherland A, Bensch S, Hellstrom T, Magg S, Wermter S. Tell Me Why You Feel That Way: Processing Compositional Dependency for Tree-LSTM Aspect Sentiment Triplet Extraction (TASTE). In: International Conference on Artificial Neural Networks. Springer; 2020. p. 660-71.

[7] Xu L, Li H, Lu W, Bing L. Position-Aware Tagging for Aspect Sentiment Triplet Extraction. arXiv preprint arXiv:201002609. 2020.

[8] Zhang C, Li Q, Song D, Wang B. A Multi-task Learning Framework for Opinion Triplet Extraction. arXiv preprint arXiv:201001512. 2020.

[9] Sun C, Huang L, Qiu X. Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence. arXiv preprint arXiv:190309588. 2019.

[10] Hoang M, Bihorac OA, Rouces J. Aspect-based sentiment analysis using bert. In: NEAL Proceedings of the 22nd Nordic Conference on Computional Linguistics (NoDaLiDa), September 30-October 2, Turku, Finland. 167. Linkoping University Electronic Press; 2019. p. 187-96.¨

[11] Li X, Bing L, Zhang W, Lam W. Exploiting BERT for end-to-end aspect-based sentiment analysis. arXiv preprint arXiv:191000883. 2019.

[12] Wang W, Pan SJ, Dahlmeier D, Xiao X. Coupled multi-layer attentions for co-extraction of aspect and opinion terms. In: Thirty-First AAAI Conference on Artificial Intelligence; 2017. .

[13] Zhao H, Huang L, Zhang R, Lu Q, et al. 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; 2020. p. 3239-48.

[14] Chen S, Liu J, Wang Y, Zhang W, Chi Z. Synchronous Double-channel Recurrent Network for AspectOpinion Pair Extraction. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics; 2020. p. 6515-24.

[15] Souza F, Nogueira R, Lotufo R. Portuguese named entity recognition using BERT-CRF. arXiv preprint arXiv:190910649. 2019.

[16] Devlin J, Chang MW, Lee K, Toutanova K. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:181004805. 2018.

[17] Lafferty J, McCallum A, Pereira FC. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. 2001.

[18] Kendall A, Gal Y, Cipolla R. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2018. p. 7482-91.

[19] Chen Z, Badrinarayanan V, Lee CY, Rabinovich A. Gradnorm: Gradient normalization for adaptive loss balancing in deep multitask networks. In: International Conference on Machine Learning. PMLR; 2018. p. 794-803.

[20] Wu Z, Ying C, Zhao F, Fan Z, Dai X, Xia R. Grid Tagging Scheme for End-to-End Fine-grained Opinion Extraction. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings; 2020. p. 2576-85.

[21] Lample G, Ballesteros M, Subramanian S, Kawakami K, Dyer C. Neural Architectures for Named Entity Recognition. In: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies; 2016. .