HITSZ-HLT at SemEval-2022 Task 10: A Span-Relation Extraction Framework for Structured Sentiment Analysis

Yihui Li, Yifan Yang, Yice Zhang, Ruifeng Xu
Harbin Institute of Technology (Shenzhen), China
20s051013@stu.hit.edu.cn, {evanyfyang,zhangyc_hit}@163.com,
xuruifeng@hit.edu.cn

Abstract

This paper describes our system that participated in the SemEval-2022 Task 10: Structured Sentiment Analysis, which aims to extract opinion tuples from texts. A full opinion tuple generally contains an opinion holder, an opinion target, the sentiment expression, and the corresponding polarity. The complex structure of the opinion tuple makes the task challenging. To address this task, we formalize it as a span-relation extraction problem and propose a two-stage extraction framework accordingly. In the first stage, we employ the span module to enumerate spans and then recognize the type of every span. In the second stage, we employ the relation module to determine the relation between spans. Our system achieves competitive results and ranks among the top-10 systems in almost subtasks.

1 Introduction

Sentiment analysis, also called opinion mining, aims to analysis people's attitudes and emotions towards specific targets, such as products, organizations, events, etc (Liu, 2012). It has become an important research field in natural language processing (Medhat et al., 2014; Hussein, 2018; Zhang et al., 2018).

Structured sentiment analysis. Barnes et al. (2021) formally defines a complete opinion as a quadruple (h, t, e, p) where h is a holder who expresses a polarity p towards a target t through a sentiment expression e. Figure 1 presents examples of opinion quadruples. On the basis of this definition, Barnes et al. (2022) formally establishes a benchmark for structured sentiment analysis. This benchmark consists of two tracks, the monolingual track and the crosslingual track, and we participate the monolingual track.

In this paper, we cast this task as a span-relation extraction problem (Jiang et al., 2020), which is a formalization that has been widely used in many

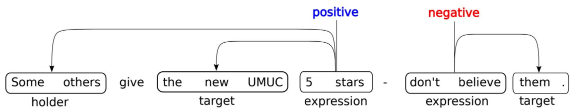

Figure 1: Examples of opinion quadruples (Barnes et al., 2021).

information extraction tasks (Eberts and Ulges, 2019; Xu et al., 2021; Lu and Ng, 2021; Li et al., 2021). With the span-relation formalization, opinion quadruple extraction is divided into two stages.

- In the first stage, we extract "meaningful" text spans and recognize their types. Specifically for this task, the type space is $\{h, t, e\}$. For those spans classified as e, we additionally detect the sentiment polarity they express.

- In the second stage, we determine the relations between spans. The relation space is set to $\{eh, et, ee, \text{none}\}$. eh and eh are used to facilitate the matching of sentiment expressions, holders, and targets during the decoding process. ee is used to deal with discontinuous sentiment expressions, which is inspired by (Li et al., 2021).

In addition, we employ span pruning (Xu et al., 2021) to reduce the computation of the second stage. Finally, opinion quadruples are decoding from the results of two stages. Our system achieves competitive results and ranks among the top-10 systems in almost subtasks.

2 Related Work

Span extraction is a fundamental method for many tasks, such as named entity recognition, aspect-level sentiment analysis, etc. This method performs element extraction by enumerating all possible spans and then determining the type of spans. Xu et al. (2017) attempts to determine the type of

Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022), pages 1406 - 1411
July 14-15, 2022 ©2022 Association for Computational Linguistics

spans by encoding all possible spans into a representation of the same size. Sohrab and Miwa (2018) also enumerate all potential spans and then use a deep network to classify them. Luan et al. (2019) leverage the coreference and relation type confidences to enhance the representation of spans. Tan et al. (2020) added the task of span boundary detection to improve the sensitivity of the model to span boundaries. This approach was able to produce higher quality candidate spans.

Span-relation extraction for sentiment tasks focuses on extracting categories of spans and relationships between spans, such as extracting relationships between entities and extracting aspect sentiment triplet. Peng et al. (2020) try to solve the aspect sentiment triplet extraction problem using a two-stage pipeline. The first stage extracts the target as well as its polarity and opinion, using the BIOES annotation method.The second stage then couples the extracted target and opinion terms to determine their paired sentiment relation. However, This method may suffer from the problem of error propagation. End-to-end methods(Wu et al., 2020; Xu et al., 2020) can extract both span and their relationships. However, previous work has usually used word-to-word interactions to predict sentiment relationships. The disadvantage of this approach is that it ignores the sentiment consistency of the entire span. The method proposed by Xu et al. (2021) can accurately enumerate all the span representations with high likelihood and then predict the sentiment relationship between them. This approach can mitigate the impact of errors in the span extraction step on subsequent relationship prediction, while it also preserves the sentiment consistency of the entire span when predicting relationships

3 Our System

Given the input text, we first obtain its contextualized representation through a pre-trained language model, BERT (Devlin et al., 2019) or RoBERTa (Liu et al., 2019). Then we input the contextualized representation into the span module and the relation module in turn to extract spans and detect relations.

3.1 Span Module

Span module roughly follows the idea of Tan et al. (2020). First we employ two binary classifiers to detect the start and end position of the "meaningful" spans respectively. Then another classifier is adopted to match the start and end positions and determine the category.

3.1.1 Start and End Prediction

Suppose $H \in \mathbb{R}^{n\times d}$ is the contextualized representation output by the language model, where n is the length of the input text. Then We calculate the probability of each position being the start or end position:

$$P_{start} = \text{sigmoid}(H \cdot W_{start}) \in \mathbb{R}^{n\times 1}, \quad (1)$$
$$P_{end} = \text{sigmoid}(H \cdot W_{end}) \in \mathbb{R}^{n\times 1}. \quad (2)$$

where $W_{start}, W_{end} \in \mathbb{R}^{d\times 1}$ are learnable parameters. Afterwards, we can decode the candidate start and end positions:

$$I_{start(>t)} = \{i \mid P^{(i)}_{start} > t, i = 1, \cdots, n\}, \quad (3)$$
$$I_{end(>t)} = \{i \mid P^{(i)}_{end} > t, i = 1, \cdots, n\}, \quad (4)$$

where threshold $t \in (0, 0.5]$ is hyper-parameter.

3.1.2 Start-End Matching and Classification

We adopt a classifier to match the start and end positions and determine the category. If the start position $i \in I_{start(>t)}$ and the end position $j \in I_{end(>t)}$ satisfy $i \leq j$, then we predict the category of span (i, j):

$$r_{ij} = [h_i; h_j; f_{width}(i, j)] \in \mathbb{R}^{3d}, \quad (5)$$
$$P^{(i,j)}_{span} = \text{softmax}(\text{FFNN}_s(r_{ij})) \in \mathbb{R}^4, \quad (6)$$

where $f_{width}(i, j) \in \mathbb{R}^d$ denotes a learnable embedding based on width $j - i$, and FFNN denotes a feed-forward neural network with non-linear activation. The span category space is $\{h, t, e, \text{invalid}\}$, where h denotes the opinion holder, t denotes the opinion target, and e denotes the sentiment expression.

For those spans classified as e, we predict its sentiment polarity additionally:

$$P^{(i,j)}_{polarity} = \text{softmax}(\text{FFNN}_p(r_{ij})) \in \mathbb{R}^3, \quad (7)$$

where the polarity space is $\{\text{POS}, \text{NEG}, \text{NEU}\}$.

3.2 Relation Module

Relation module aims to determine the relations between spans. For a span pair, we first construct a relation representation based on the span representations and then feed it into a relation classifier. Notice that we employ span pruning (Xu et al., 2021) to reduce the computation.

Language	Pretrained Model
English	roberta-large (Liu et al., 2019)
Spanish	BSC-TeMU/roberta-base-bne (Gutiérrez-Fandiño et al., 2021)
Norwegian	pere/norwegian-roberta-base
Basque	ixa-ehu/berteus-base-cased (Agerri et al., 2020)
Catalan	BSC-TeMU/roberta-base-ca (Armengol-Estapé et al., 2021)

Table 1: Pretrained language model for 5 different languages.

3.2.1 Span Pruning

Considering the large number of the predicted spans, it is not computationally practical to consider all possible pairwise relations. Following Xu et al. (2021), we prune spans in the relation classification stage. The holder, target, and sentiment expression candidates are selected based on the scores of the mention types for each span:

$$\Phi^{(i,j)}_{holder} = P^{(i,j)}_{span}(m = h), \quad (8)$$

$$\Phi^{(i,j)}_{target} = P^{(i,j)}_{span}(m = t), \quad (9)$$

$$\Phi^{(i,j)}_{expression} = P^{(i,j)}_{span}(m = e). \quad (10)$$

We use the mention scores Φ_{source}, Φ_{target} and $\Phi_{expression}$ to select the top k candidates and obtain the holder candidate pool S^h, the target candidate pool S^t, and the sentiment expression candidate pool S^e, respectively. The value of k is related to the length of the sentence n:

$$k = \max(n \cdot z, k_{min}), \quad (11)$$

where z, k_{min} are hyper-parameters.

3.3 Datasets

	Language	Domain	Train	Dev	Test
MPQA	English	news	5873	2063	2112
DS_{Unis}	English	e-commerce	2252	232	318
$\text{OpenNER}_{\text{EN}}$	English	hotel	1745	250	499
$\text{OpenNER}_{\text{ES}}$	Spanish	hotel	1439	206	410
$\text{NoReC}_{\text{Fine}}$	Norwegian	multi-domain	8634	1531	1272
$\text{MultiB}_{\text{EU}}$	Basque	hotel	1064	152	305
$\text{MultiB}_{\text{CA}}$	Catalan	hotel	1174	168	335

Table 2: Data statistics.

3.3.1 Relation Classification

For most datasets, we only detect two relations, `expression-holder` and `expression-holder`. For datasets with discontinuous sentiment expressions, we detect `expression-expression` relation additionally. We obtain the candidate pair representation by coupling each expression candidate representation $s^e_{a,b}$ with the other candidate representation. For an expression candidate $(a, b) \in S^e$ and a holder candidate $(c, d) \in S^h$, their pair representation is:

$$g^{e,h}_{(a,b),(c,d)} = [r_{a,b}; r_{c,d}; f_{distance}(a, b, c, d) \\ f_{context}(a, b, c, d); f_{type}(e); f_{type}(h)]$$

where $f_{distance} \in R^d$ denotes a learnable embedding based distance $min(|b-c|, |a-d|)$, $f_{context} \in R^d$ is obtained by performing max-pooling operation on the context between the two spans, and f_{type} is a learnable embedding for indicating the span type. We construct $g^{e,t}$, $g^{e,e}$ in a similar way.

Then we input the pair representation to a feed-forward neural network to determine the sentiment relation:

$$P^{((a,b),(c,d))}_{relation} = \text{softmax}(\text{FFNN}_r(g^{e,h}_{(a,b),(c,d)})),$$

where the relation space is $\{eh, et, ee, \text{none}\}$.

3.4 Training

During training, we utilize the cross-entropy function to calculate the loss of start & end prediction, span classification(SC), polarity classification(PC), and relation classification(RC). The overall optimization objective is to minimize the summation of these losses:

$$\mathcal{L} = \mathcal{L}_S + \mathcal{L}_E + \mathcal{L}_{SC} + \mathcal{L}_{PC} + \mathcal{L}_{RC}. \quad (12)$$

3.5 Sentiment Structure Decoding

We first decode the sentiment expressions and their sentiment polarities from the results of the span module. Then we obtain the holder candidate pool and the target candidate pool by span pruning. For each sentiment expression, we determine whether it has a relation with each holder candidate and target candidate. Finally, the opinion quadruplets are produced based on the result of the relation classification. In addition, for discontinuous sentiment expressions, sentiment expressions are merged according to the relation between sentiment expressions.

Model	MPQA	DS_{Unis}	$OpenNER_{EN}$	$OpenNER_{ES}$	$NoReC_{Fine}$	$MultiB_{EU}$	$MultiB_{CA}$
head first	17.40	25.00	-	-	29.50	56.80	54.70
head final	18.80	26.50	-	-	**31.20**	53.70	54.70
Span-Relation	**35.00**(9)	**44.90**(4)	**70.30**(8)	**64.20**(10)	21.30(21)	**63.90**(10)	**63.50**(12)

Table 3: Results on the test dataset (Sentiment Graph F_1, %).

4 Experiments

The monolingual track (Barnes et al., 2022) provides 7 structured sentiment datasets in five languages. Their statistics are listed on Table 2.

It is worth noting that there are discontinuous spans in the $NoReC_{Fine}$ and DS_{Unis} datasets. For example, in "*It looks again like UMUC will do anything for money*", "*looks again*" and "*do anything*" are annotated as the same sentiment expression.

4.1 Experiment Settings

We use BERT or RoBERTa as the text encoders. Since this task has datasets in different languages, different pre-training models are used for different language, which is detailed in Table 1.

We used Adam as our optimizer. The maximum number of epochs is set to 15, z is set to 0.3, and $k_{\min}$ is set to 5. We train our model on the training set and keep the model that performs best on the validation set. We evaluate our model on Sentiment Graph F_1 (Barnes et al., 2021) and compare our model with sentiment graph approaches (Head-first/Head-final) (Barnes et al., 2021).

4.2 Main Results

The comparison results of opinion quadruple extraction are listed in Table 3. According to these results, our approach achieves better performance on most datasets than baselines, especially on MPQA exceeding baseline by 16.2%. This demonstrates the effectiveness of our approach for opinion quadruple extraction.

4.3 Ablation Study

Model	MPQA	DS_{Unis}	$OpenNER_{EN}$
Full Model	40.67	40.04	72.38
w/o f_{width}	37.50	37.40	71.14
w/o $f_{distance}$	38.83	36.42	71.46
w/o $f_{context}$	37.54	39.47	69.39

Table 4: Ablation results on the dev dataset.

We conduct an ablation study to examine the impact of some components in the proposed model and list the results in Table 4. It can be observed that the removal of width embedding, position embedding, and context all degrade the performance, indicating their necessity.

Model	$OpenNER_{ES}$	$NoReC_{Fine}$	$MultiB_{EU}$	$MultiB_{CA}$
	62.40	23.26	61.53	54.26
w mBERT	61.62	36.22	57.17	63.35

Table 5: Effect of mBERT representations.

In addition, we also compare the performance of the multilingual pre-trained model mBERT(bert-base-multilingual-cased)(Devlin et al., 2019) for this task. To this end, we compare the experimental performance of monolingual pre-trained models with mBERT on minor language datasets and list the results in Table 5. It can be observed that mBERT achieves similar performance to the monolingual pre-trained model for most minor languages. In addition, for the Norwegian and Catalan datasets, the performance of the models with mBERT improves considerably, which may be due to the lack of corpus in these two languages when training the monolingual pre-trained models.

5 Conclusions

This paper describes our system for structured sentiment analysis. We formalize the task as a span-relation extraction problem and propose a two-stage extraction approach, which consists of a span module and a relation module. Experimental results demonstrate the effectiveness of our approach.

References

Rodrigo Agerri, Iñaki San Vicente, Jon Ander Campos, Ander Barrena, Xabier Saralegi, Aitor Soroa, and Eneko Agirre. 2020. Give your text representation models some love: the case for basque. In *Proceedings of the 12th International Conference on Language Resources and Evaluation*.

Jordi Armengol-Estapé, Casimiro Pio Carrino, Carlos Rodriguez-Penagos, Ona de Gibert Bonet, Carme Armentano-Oller, Aitor Gonzalez-Agirre, Maite

Melero, and Marta Villegas. 2021. Are multilingual models the best choice for moderately under-resourced languages? A comprehensive assessment for Catalan. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 4933–4946, Online. Association for Computational Linguistics.

Jeremy Barnes, Robin Kurtz, Stephan Oepen, Lilja Øvrelid, and Erik Velldal. 2021. Structured sentiment analysis as dependency graph parsing. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 3387–3402, Online. Association for Computational Linguistics.

Jeremy Barnes, Oberländer-Laura Ana Maria Kutuzov, Andrey and, Enrica Troiano, Jan Buchmann, Rodrigo Agerri, Lilja Øvrelid, Erik Velldal, and Stephan Oepen. 2022. SemEval-2022 task 10: Structured sentiment analysis. In *Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022)*, Seattle. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *NAACL-HLT (1)*.

Markus Eberts and Adrian Ulges. 2019. Span-based joint entity and relation extraction with transformer pre-training. *arXiv preprint arXiv:1909.07755*.

Asier Gutiérrez-Fandiño, Jordi Armengol-Estapé, Marc Pàmies, Joan Llop-Palao, Joaquín Silveira-Ocampo, Casimiro Pio Carrino, Aitor Gonzalez-Agirre, Carme Armentano-Oller, Carlos Rodriguez-Penagos, and Marta Villegas. 2021. Spanish language models.

Doaa Mohey El-Din Mohamed Hussein. 2018. A survey on sentiment analysis challenges. *Journal of King Saud University-Engineering Sciences*, 30(4):330–338.

Zhengbao Jiang, Wei Xu, Jun Araki, and Graham Neubig. 2020. Generalizing natural language analysis through span-relation representations. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 2120–2133, Online. Association for Computational Linguistics.

Fei Li, ZhiChao Lin, Meishan Zhang, and Donghong Ji. 2021. A span-based model for joint overlapped and discontinuous named entity recognition. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 4814–4828, Online. Association for Computational Linguistics.

Bing Liu. 2012. Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1):1–167.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized BERT pretraining approach. *CoRR*, abs/1907.11692.

Jing Lu and Vincent Ng. 2021. Span-based event coreference resolution. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 13489–13497.

Yi Luan, Dave Wadden, Luheng He, Amy Shah, Mari Ostendorf, and Hannaneh Hajishirzi. 2019. A general framework for information extraction using dynamic span graphs. In *NAACL-HLT (1)*.

Walaa Medhat, Ahmed Hassan, and Hoda Korashy. 2014. Sentiment analysis algorithms and applications: A survey. *Ain Shams engineering journal*, 5(4):1093–1113.

Haiyun Peng, Lu Xu, Lidong Bing, Fei Huang, Wei Lu, and Luo Si. 2020. Knowing what, how and why: A near complete solution for aspect-based sentiment analysis. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 8600–8607.

Mohammad Golam Sohrab and Makoto Miwa. 2018. Deep exhaustive model for nested named entity recognition. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2843–2849.

Chuanqi Tan, Wei Qiu, Mosha Chen, Rui Wang, and Fei Huang. 2020. Boundary enhanced neural span classification for nested named entity recognition. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 9016–9023.

Zhen Wu, Chengcan Ying, Fei Zhao, Zhifang Fan, Xinyu Dai, and Rui Xia. 2020. Grid tagging scheme for aspect-oriented fine-grained opinion extraction. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 2576–2585.

Lu Xu, Yew Ken Chia, and Lidong Bing. 2021. Learning span-level interactions for aspect sentiment triplet extraction. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 4755–4766.

Lu Xu, Hao Li, Wei Lu, and Lidong Bing. 2020. Position-aware tagging for aspect sentiment triplet extraction. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2339–2349.

Mingbin Xu, Hui Jiang, and Sedtawut Watcharawittayakul. 2017. A local detection approach for named entity recognition and mention detection. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1237–1247.

Lei Zhang, Shuai Wang, and Bing Liu. 2018. Deep learning for sentiment analysis: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(4):e1253.