Appendix for „Exploiting Phrase Interrelations in Span-level Neural Approaches for Aspect Sentiment Triplet Extraction”

Iwo Naglik and Mateusz Lango[0000−0003−2881−5642]

Poznan University of Technology, Institute of Computing Science, Poznań, Poland
inaglik.put.poznan@gmail.com, mlango@cs.put.poznan.pl

1 Decoding algorithm

This section will present the pseudo-code of the decoding algorithm. This algorithm is used to obtain the final triples from the prediction matrix created.

The function $agree\_predictions(matrix)$ unifies the predictions of the model. It makes modifications to the resulting probabilities. Because of the way the matrix is formulated, there can only be predictions on the diagonal that refer to the phrase (aspect phrase, opinion phrase, or invalid), not the relationship between the selected phrases. On the other hand, above the diagonal are the relations of the selected phrases in the form of an assigned sentiment, if a pair of phrases should be combined, or information about an invalid pairing. Therefore, the diagonal of the matrix can be made consistent by zeroing out the probabilities of assigning sentiment labels in its elements and distributing the remaining results accordingly to obtain the correct probability distribution. The same can be done in the elements above the diagonal by zeroing the probabilities of assigning an aspect or opinion phrase class and aligning the remaining results accordingly.

The rest of the algorithm is presented in the main part of the article and can be described by the algorithm 1.

Algorithm 1 Decoding algorithm

```
// A - Aspect class, O - Opinion class, I - Invalid class //
agree_predictions(matrix)
results = {}
for diag_el in matrix_diagonal do
    if diag_el not in (A, O) then
        continue
    end if
    for col_el in {0...index(diag_el)} do
        relation = matrix[index(col_el), index(diag_el)]
        if (col_el not in (A, O)) or (col_el = diag_el) or (relation = I) then
            continue
        end if
        results.add([index(col_el), index(diag_el), relation])
    end for
end for
```

2 Model architecture

This section will introduce the various elements of the model and briefly discuss the process of training the entire architecture.

Algorithm 2 Full architecture training process

```
// mlm - Masked Language Model
span_constructor.unfreeze(), span_filter.unfreeze(), mlm.unfreeze()
triplets.freeze()
train(epochs_num=n)
triplets.unfreeze()
span_constructor.freeze(), span_filter.freeze(), mlm.freeze()
train(epochs_num=m)                                          ▷ m < n
span_constructor.unfreeze(), span_filter.unfreeze(), mlm.unfreeze(),
triplets.unfreeze()
train(epochs_num=k)                                          ▷ k ≫ n, m
```

2.1 Span Constructor block

Algorithm 3 Span Constructor block

```
linear_layer1(input_dim, input_dim // 2)
linear_layer2(input_dim // 2, 5) // five - number of BIO possible classes
CRF(num_tags=5)
```

2.2 Span Filter block

Algorithm 4 Span Filter block

```
linear_layer1(input_dim, 300)
ReLU()
Dropout(0.1)
linear_layer2(300, 100)
ReLU()
Dropout(0.1)
final_layer(100, 1)
Sigmoid()
```

2.3 Triplet block

Algorithm 5 Triplet block

```
batch_norm()
linear_layer1(input_dim, 300)
ReLU()
Dropout(0.1)
linear_layer2(300, 100)
ReLU()
Dropout(0.1)
linear_layer3(100, 100)
ReLU()
Dropout(0.1)
batch_norm()
final_layer(100, 6) // Number of possible classes
FC_CRF(filter_size=3, smoothness_theta=0.85, smoothness_weight=1)
Softmax()
```

3 Detailed results

This section will present the statistics of the results obtained and compare the predictions of the model in comparison with the ground truth. The results presented are those obtained by a single, arbitrarily chosen model (they are not an average of several runs). They are intended to indicate potential areas of improvement and difficulties.

3.1 General statistics

Obtained statistics for selected datasets. In the left column are original statistics, and in the right column statistics obtained by the model. Additionally - a summary of results for negation phrases occurring in datasets

	original_labels	model_outputs
Number of sentences	328	328
Mean sentence length	15.7652	15.7652
Number of opinion phrases	470	444
Number of aspect phrases	463	442
Number of triplets	543	536
Mean length of opinion phrases	1.1383	1.0923
Mean length of aspect phrases	1.4017	1.3416
Number of positive sentiment	364	357
Number of neutral sentiment	63	26
Number of negative sentiment	116	153
Number of one-to-many relations (opinions)	56	76
Number of one-to-many relations (aspects)	67	85
Number of opinions with length $= 1$	405	403
Number of aspects with length $= 1$	277	292
Number of opinions with length > 1	65	41
Number of aspects with length > 1	186	150
Number of triplets where length of each span $= 1$	291	334
Number of triplets where aspect span length > 1 and opinion span length $= 1$	183	156
Number of triplets where opinion span length > 1 and aspect span length $= 1$	32	18
Number of triplets where at least one span length > 1	252	202

Table 1. Quantitative characteristics of the results for the 14lap dataset

		not	n't
original_labels	Number of specific phrases	45	20
original_labels	Number of specific phrases in aspect span	0	0
original_labels	Number of specific phrases in opinion span	32	5
model_outputs	Number of specific phrases	45	20
model_outputs	Number of specific phrases in aspect span	0	0
model_outputs	Number of specific phrases in opinion span	25	5

Table 2. Summary of results for negation phrases occurring in the 14lap dataset

	original_labels	model_outputs
Number of sentences	492	492
Mean sentence length	16.3435	16.3435
Number of opinion phrases	845	802
Number of aspect phrases	848	788
Number of triplets	994	935
Mean length of opinion phrases	1.1018	1.0661
Mean length of aspect phrases	1.2712	1.2259
Number of positive sentiment	773	764
Number of neutral sentiment	66	29
Number of negative sentiment	155	142
Number of one-to-many relations (opinions)	93	97
Number of one-to-many relations (aspects)	128	131
Number of opinions with length = 1	759	749
Number of aspects with length = 1	618	615
Number of opinions with length > 1	86	53
Number of aspects with length > 1	230	173
Number of triplets where length of each span = 1	657	681
Number of triplets where aspect span length > 1 and opinion span length = 1	246	191
Number of triplets where opinion span length > 1 and aspect span length = 1	68	50
Number of triplets where at least one span length > 1	337	254

Table 3. Quantitative characteristics of the results for the 14res dataset

		not	n't
original_labels	Number of specific phrases	46	22
original_labels	Number of specific phrases in aspect span	0	0
original_labels	Number of specific phrases in opinion span	14	4
model_outputs	Number of specific phrases	46	22
model_outputs	Number of specific phrases in aspect span	0	0
model_outputs	Number of specific phrases in opinion span	10	3

Table 4. Summary of results for negation phrases occurring in the 14res dataset

	original_labels	model_outputs
Number of sentences	322	322
Mean sentence length	15.6273	15.6273
Number of opinion phrases	456	430
Number of aspect phrases	432	386
Number of triplets	485	468
Mean length of opinion phrases	1.1382	1.1209
Mean length of aspect phrases	1.2917	1.2824
Number of positive sentiment	317	334
Number of neutral sentiment	25	18
Number of negative sentiment	143	116
Number of one-to-many relations (opinions)	26	34
Number of one-to-many relations (aspects)	47	72
Number of opinions with length = 1	393	380
Number of aspects with length = 1	306	277
Number of opinions with length > 1	63	50
Number of aspects with length > 1	126	109
Number of triplets where length of each span = 1	297	300
Number of triplets where aspect span length > 1 and opinion span length = 1	124	118
Number of triplets where opinion span length > 1 and aspect span length = 1	51	40
Number of triplets where at least one span length > 1	188	168

Table 5. Quantitative characteristics of the results for the 15res dataset

		not	n't
original_labels	Number of specific phrases	36	17
original_labels	Number of specific phrases in aspect span	0	0
original_labels	Number of specific phrases in opinion span	16	4
model_outputs	Number of specific phrases	36	17
model_outputs	Number of specific phrases in aspect span	0	0
model_outputs	Number of specific phrases in opinion span	16	4

Table 6. Summary of results for negation phrases occurring in the 15res dataset

	original_labels	model_outputs
Number of sentences	326	326
Mean sentence length	14.6963	14.6963
Number of opinion phrases	470	467
Number of aspect phrases	452	451
Number of triplets	514	536
Mean length of opinion phrases	1.1106	1.0621
Mean length of aspect phrases	1.2655	1.2705
Number of positive sentiment	407	458
Number of neutral sentiment	29	4
Number of negative sentiment	78	74
Number of one-to-many relations (opinions)	27	40
Number of one-to-many relations (aspects)	56	74
Number of opinions with length = 1	418	438
Number of aspects with length = 1	332	331
Number of opinions with length > 1	52	29
Number of aspects with length > 1	120	120
Number of triplets where length of each span = 1	344	379
Number of triplets where aspect span length > 1 and opinion span length = 1	116	126
Number of triplets where opinion span length > 1 and aspect span length = 1	41	24
Number of triplets where at least one span length > 1	170	157

Table 7. Quantitative characteristics of the results for the 16res dataset

		not	n't
original_labels	Number of specific phrases	23	17
original_labels	Number of specific phrases in aspect span	0	0
original_labels	Number of specific phrases in opinion span	8	8
model_outputs	Number of specific phrases	23	17
model_outputs	Number of specific phrases in aspect span	0	0
model_outputs	Number of specific phrases in opinion span	7	1

Table 8. Summary of results for negation phrases occurring in the 16res dataset

It is worth paying special attention at this point, to the problems of the model in including words indicating a negation in opinion phrases. This could result in the assigned sentiment being incorrect due to the model's failure to handle the aforementioned negation. In the future, it is worth paying special attention to strengthening this element, as this could translate into increases in the quality of the model's performance.

Also, a key element is an observation that the model has difficulty in predicting the sentiment of a neutral class. This may be due to the fact that there were far fewer instances representing this class in the data sets and it was more difficult to learn the differences between the neutral class and the extreme classes. The model was more likely to make mistakes in classifying the neutral class and assigning it the positive class label. Also, for people, distinguishing the sentiment of neutral from positive can be a problem, as the boundaries between the two intersect.

3.2 Quantitative and qualitative analysis of the phrase extraction process

To further show how drastically the number of extracted and analyzed phrases was reduced, a deeper analysis was performed. The following tables contain information about the number and effectiveness of extracted phrases in the context of each model. The row of the table, „all possible phrases", contains a model that, as in [1] work, creates all possible phrases with a sliding window with a limitation on the maximum window length of 5.

Precision, sensitivity, and F1 have the same interpretation as in the final analysis of the results, focusing only on the accuracy of phrase extraction and not on the correct pairing and assignment of sentiment. The number of phrases extracted indicates how many phrases were extracted through the corresponding model responsible for generating spans for analysis. This indicates the mean number of elements that are further analyzed in the context of the pairing task and the assignment of the appropriate sentiment (note that some phrases are incorrect and should be omitted - possible errors at this stage). The number of correct phrases indicates the number of correct phrases in the dataset.

		Precision	Recall	F1	Extracted spans	Total correct spans
14lap	our model	0.69	0.69	0.69	932.4	937.0
14lap	all possible phrases	0.04	0.997	0.07	24167	937
14res	our model	0.85	0.88	0.87	1765.2	1701.0
14res	all possible phrases	0.04	1	0.09	37917	1701
15res	our model	0.80	0.83	0.82	923.2	893.0
15res	all possible phrases	0.04	0.98	0.07	23586	893
16res	our model	0.83	0.89	0.86	988.2	926.0
16res	all possible phrases	0.04	0.98	0.08	22063	926

Table 9. Analysis of the phrase extraction process.

As can be seen, our model produces less number of phrases. One can conclude that, indeed, the reduction of the analyzed ranges contributes to obtaining better and more consistent results when using the relation matrix module. It can be inferred from this that focusing on smartly reducing the generated phrases could lead to constructing better final results.

3.3 Error analysis

This section will analyze the possible causes of errors made by the model.

Number of bad pairing	Bad sentiment	Bad (opposite) sentiment
10	48	10

Table 10. Analysis of errors in attribution of sentiment. Dataset: 14lap

Number of bad pairing	Bad sentiment	Bad (opposite) sentiment
6	34	3

Table 11. Analysis of errors in attribution of sentiment. Dataset: 14res

Number of bad pairing	Bad sentiment	Bad (opposite) sentiment
6	28	20

Table 12. Analysis of errors in attribution of sentiment. Dataset: 15res

Number of bad pairing	Bad sentiment	Bad (opposite) sentiment
1	24	9

Table 13. Analysis of errors in attribution of sentiment. Dataset: 16res

This analysis allows us to see that the model was much more likely to make errors related to the wrong classification of sentiment than to the bad pairing of correct phrases. As has been noted earlier, the problem of wrong sentiment assignment is much greater in the case of neutral class mistakes than in the case of extreme sentiment classes.

3.4 Error analysis of phrase generation

This section will show the model's problems when generating phrases. Phrases will be listed, along with the counts (how many times the situation occurred) that were not included in the created phrases or were excessively included in the resulting ranges.

	0
Not included words: much	1
Not included words: finger	1
Not included words: product	1
Not included words: customize	1
Not included words: overall	2
Not included words: for	1
Not included words: stuff	1
Not included words: of	2
Not included words: lighted	2
Not included words: Mac	2
Not included words: up	2
Not included words: bluetooth	1
Not included words: integrate	1
Not included words: (	4
Not included words: utterly	1
Not included words: anodized	4
Not included words: 2.9ghz	1
Not included words: compact	1
Not included words: keys	1
Not included words: provided	1
Not included words: built-in	1
Not included words: have	1
Over included words: restrictions	1
Over included words: of	2
Over included words: devices	1
Over included words: i7	1
Over included words: power	1
Over included words: jaw	1

Table 14. Analysis of containing elements in phrases. Dataset: 14lap

	0
Not included words: options	2
Not included words: entree-sized	1
Not included words: spicy	3
Not included words: fresh	2
Not included words: Creamy	2
Not included words: bistro	1
Not included words: with	5
Not included words: macadamia-crusted	1
Not included words: sweet	2
Not included words: 's	6
Not included words: mouth	1
Not included words: afternoon	1
Not included words: organic	4
Not included words: garnished	2
Not included words: equally	1
Not included words: casual	1
Not included words: orange	1
Not included words: experience	3
Not included words: AND	1
Not included words: artifical	2
Not included words: Tuna	1
Not included words: boutique	2
Not included words: and	1
Not included words: (	5
Not included words: of	3
Not included words: priced	2
Not included words: new	2
Not included words: bi-level	1
Not included words: long	2
Not included words: soft	2
Not included words: big	1
Not included words: than	1
Not included words: homemade	2
Over included words: roll	1
Over included words: spicy	1
Over included words: concoctions	1
Over included words: entree	1
Over included words: short	2
Over included words: beef	2
Over included words: burgers	4
Over included words: fell	1
Over included words: Irish	4
Over included words: meat	2
Over included words: of	2
Over included words: scene	4
Over included words: appetizer	1
Over included words: items	4
Over included words: sizzling	1
Over included words: not	1
Over included words: crab	2

Table 15. Analysis of containing elements in phrases. Dataset: 14res

	0
Not included words: in	1
Not included words: and	1
Not included words: expect	1
Not included words: modern	2
Not included words: white	1
Not included words: our	1
Not included words: alla	1
Not included words: of	1
Not included words: The	1
Not included words: as	1
Not included words: with	1
Not included words: for	1
Over included words: food	1
Over included words: and	1
Over included words: white	1
Over included words: mean	1
Over included words: quality	1
Over included words: experience	2
Over included words: Seasons	1
Over included words: tap	2
Over included words: at	1

Table 16. Analysis of containing elements in phrases. Dataset: 15res

	0
Not included words: on	1
Not included words: and	1
Not included words: open	1
Not included words: quality	1
Not included words: setting	1
Not included words: even	1
Not included words: Japanese	1
Not included words: brow	1
Not included words: cooked	1
Not included words: dinners	1
Over included words: Neighborhood	1
Over included words: ,	1
Over included words: smoking	1
Over included words: suggestions	1
Over included words: prawns	4
Over included words: frozen	1

Table 17. Analysis of containing elements in phrases. Dataset: 16res

References

1. Xu, L., Chia, Y.K., Bing, L.: Learning span-level interactions for aspect sentiment triplet extraction. In: Proceedings of the 59th Annual Meeting of the ACL and the 11th IJCNLP (Volume 1: Long Papers). pp. 4755–4766. Association for Computational Linguistics (2021)