Improving Aspect-Based Sentiment with End-to-End Semantic Role Labeling Model

Pavel Přibáň[*] **and Ondřej Pražák**[*]

University of West Bohemia, Faculty of Applied Sciences, Czech Republic
[1]Department of Computer Science and Engineering,
[1]NTIS – New Technologies for the Information Society,
{pribanp,ondfa}@kiv.zcu.cz
http://nlp.kiv.zcu.cz

Abstract

This paper presents a series of approaches aimed at enhancing the performance of Aspect-Based Sentiment Analysis (ABSA) by utilizing extracted semantic information from a Semantic Role Labeling (SRL) model. We propose a novel end-to-end Semantic Role Labeling model that effectively captures most of the structured semantic information within the Transformer hidden state. We believe that this end-to-end model is well-suited for our newly proposed models that incorporate semantic information. We evaluate the proposed models in two languages, English and Czech, employing ELECTRA-small models. Our combined models improve ABSA performance in both languages. Moreover, we achieved new state-of-the-art results on the Czech ABSA.

1 Introduction

In recent years, the pre-trained BERT-like models based on the Transformer (Vaswani et al., 2017) architecture demonstrated their performance superiority across various natural language processing (NLP) tasks. In this paper, we study the possibility of a combination of two seemingly unrelated NLP tasks: Aspect-Based Sentiment Analysis (ABSA) and Semantic Role Labeling (SRL). We believe that the structured semantic information of a sentence extracted from an SRL model can enhance the performance of an ABSA model. We investigate our assumption on the ELECTRA (Clark et al., 2020) model architecture since it is a lighter and smaller alternative to the popular and commonly used models such as BERT (Devlin et al., 2019) or RoBERTa (Liu et al., 2019). Because the ELECTRA model is smaller in terms of the number of parameters, it does require less GPU memory and time to be fine-tuned.

Sentiment analysis (SA) is an essential part of NLP. The most prevalent SA task is the *Sentiment Classification*, where the objective is to classify a text fragment (e.g., sentence or review) as *positive* or *negative*, eventually as *neutral*. In this type of task, we assume that there is only one opinion in the text. In reality, as illustrated in Figure 1, this assumption often does not hold true (Liu, 2012).

"*The burger was excellent but the waitress was unpleasant*"
　　CE ⇒ food, service
　　CP ⇒ food:*positive*, service:*negative*

Figure 1: Example of CE and CP subtasks of ABSA.

Aspect-Based Sentiment Analysis (Liu, 2012; Pontiki et al., 2014) focuses on detecting aspects (e.g., food or service in the restaurant reviews domain) and determining their polarity, enabling more detailed analysis and understating of the expressed sentiment. As shown by Pontiki et al. (2014), the ABSA task can be further divided into four subtasks: *Aspect term extraction* (TE), *Aspect term polarity* (TP), *Aspect category extraction* (CE), and *Aspect category polarity* (CP).

We aim at the CE and CP subtasks,[1] and we treat them as a single classification task, see Section 3.2. As depicted in Figure 1, the goal of the CE subtask is to detect a set of aspect categories within a given sentence, i.e., for a given text $S = \{w_1, w_2, \ldots w_n\}$ assign set $M = \{a_1, a_2, \ldots, a_m\}$ of m aspect categories, where $m \in [0, k]$, $M \subset A$ and A is a set of k predefined aspect categories $A = \{a_1, a_2, \ldots, a_k\}$. The goal of CP is to assign one of the predefined polarity labels p for each of the given (or predicted) aspect categories of the set M for the given text S, where $p \in P = \{positive, negative, neutral\}$.

The Semantic Role Labeling task (Gildea and Jurafsky, 2002) belongs among shallow semantic

---

[*]Equal contribution.
[1]See (Pontiki et al., 2014) for a detailed description of all the subtasks.

parsing techniques. The SRL goal is to identify and categorize semantic relationships or *semantic roles* of given *predicates*. Verbs, such as "believe" or "cook", are natural predicates, but certain nouns are also accepted as predicates. The simplified definition of semantic roles is that semantic roles are abstractions of predicate arguments. For example, the semantic roles for "believe" can be *Agent* (a believer) and *Theme* (a statement) and for "cook" *Agent* (a chef), *Patient* (a food), *Instrument* (a device for cooking) – see examples in Figure 2. The theory of predicates and their roles is very well established in several linguistic resources such as PropBank (Palmer et al., 2005) or FrameNet (Baker et al., 1998).

(1) [He]$_{\text{AGENT}|\text{A0}}$ <u>believes</u> [in what he plays] $_{\text{THEME}|\text{A1}}$.

(2) Can [you] $_{\text{AGENT}|\text{A0}}$ <u>cook</u> [the dinner] $_{\text{PATIENT}|\text{A1}}$?

Figure 2: Examples of SRL annotations.

In this work, we introduce a novel end-to-end SRL model that offers enhanced compatibility with other NLP tasks. Unlike other BERT-based models (Shi and Lin, 2019; Papay et al., 2021), our proposed approach integrates the complete semantic information into the hidden state of the Transformer. This end-to-end SRL model is particularly well-suited for combination with the Aspect-Based Sentiment Analysis task, as it encapsulates the entire predicate-argument structure of the sentence within a single hidden state, in contrast to the approach of (Shi and Lin, 2019), which encodes each predicate separately and requires gold predicates on input. Our model, on the other hand, only requires the input text.

We assume that leveraging the syntax and semantic information extracted from SRL can significantly enhance the performance of the aspect category polarity subtask. This assumption is grounded in the notion that the SRL information has the potential to unveil valuable and pertinent relations between entities within a given sentence, which play a crucial role in accurate aspect category polarity predictions. This holds particularly true for longer and more complex sentences, where a broader contextual understanding becomes essential. For a concrete illustration, please refer to Appendix B.

To combine the SRL and ABSA models effectively, we propose three different approaches. Through their integration, we demonstrate performance improvements on the ABSA task for both English and Czech languages, employing

ELECTRA-small models. Moreover, we achieved new state-of-the-art (SotA) results on the Czech ABSA task. We publicly release our source codes[2].

2 Related Work

The early studies (Hu and Liu, 2004; Ganu et al., 2009; Kiritchenko et al., 2014; Hercig et al., 2016) focusing on the English ABSA task relied on word n-grams, lexicons, and other feature extraction techniques in combination with supervised machine learning algorithm such as support vector machine classifiers. These approaches were surpassed by deep neural network (DNN) models (Tang et al., 2016; Ma et al., 2017; Chen et al., 2017; Fan et al., 2018) that typically employed recurrent neural network e.g., Long Short-Term Memory (LSTM) (Hochreiter and Schmidhuber, 1997).

Recently, the BERT-like models were successfully applied to the ABSA task. Sun et al. (2019) solve the CE and CP subtasks at once by introducing auxiliary sentences and transforming the problem to a sentence-pair classification task. Xu et al. (2019) and Rietzler et al. (2020) improved results by pre-training the model on the task domain data. Liu et al. (2021) treated the ABSA task as a text generation task outperforming the previous SotA results. (Zhang et al., 2019; Liang et al., 2022) employed graph convolutional networks. Another related work can be found in (Li et al., 2020).

In (Sido et al., 2021; Přibáň and Steinberger, 2021; Lehečka et al., 2020; Přibáň and Steinberger, 2022) the BERT-like models were used for sentiment classification and subjectivity classification, to the best of our knowledge, there is no application of BERT-like models for ABSA in the Czech language. Steinberger et al. (2014) introduced the first Czech ABSA dataset from the restaurant reviews domain. They used a Maximum Entropy classifier and Conditional Random Fields for their baselines. Hercig et al. (2016) extended this dataset and improved the baseline by adding semantic features. Lenc and Hercig (2016) applied a convolutional neural network for the CP task and RNN for the CP task to the dataset from Hercig et al. (2016).

The pioneered approaches of the SRL (Gildea and Jurafsky, 2002) task used standard feature engineering methods (Moschitti et al., 2008). Since SRL is closely bounded with syntax, adding syntactic information is very helpful. In 2008 CoNLL

---

[2]https://github.com/pauli31/srl-aspect-based-sentiment

shared task (Surdeanu et al., 2008) syntax-based SRL task was proposed.

In more recent years (with DNNs), the attention was drawn back to standard span-based SRL, where we form SRL as (linear) tagging. Many approaches are based on LSTMs (He et al., 2017). Later, Tan et al. (2018), inspired by the Transformer, proposed a self-attention-based model.

Several end-to-end models for all SRL subtasks were also introduced. He et al. (2018) abandon the BIO tagging scheme, and they are rather predicting predicate-argument span tuples by searching through the possible combinations. They use a multi-layer bi-LSTM to produce contextualized representations of predicates and argument spans. The most recent approaches use BERT-like pretrained models. Shi and Lin (2019) proposed a simple BERT approach for argument identification and classification. This means, in their setting, the gold predicates are known. Papay et al. (2021) propose regular-constrained conditional random fields (CRF) decoding on top of the same model. There are many other complex deep models (Zhang et al., 2021; Wang et al., 2021)

For our experiments, we need an end-to-end SRL model which encodes most of the information in the Transformer's hidden state. However, to the best of our knowledge, there is no such model. As a result, we introduce our end-to-end model later in this paper to fulfil this need.

Various approaches have been made to enhance one task through the integration of another, usually using multi-task learning techniques. Hashimoto et al. (2016) proposed a joint model for learning the whole NLP stack (POS tagging, chunking, parsing, semantic relatedness, entailment). They train a single model for all tasks in a sequence (chunking after POS tagging etc.). At each layer (for each task), they use regularization on the difference from previous layer weights. They show that the tasks help each other significantly.

Li et al. (2021) use dependency neighbourhood prediction and part-of-speech tagging as auxiliary tasks for ABSA. They introduced the new dependency neighbourhood prediction task to utilize the syntactic dependency information to improve the performance of the sentiment classification task. They train the auxiliary tasks together with the main sentiment classification task. The task classifies each token as either in the dependency neighbourhood or not. The dependency neighbourhood

for a given token in a sentence is defined as the tokens in the sentence that are linked to the given token through, at most, n-hop dependency relations. Zhang et al. (2020) pretrain BERT model on semantic role labeling task and show, that the pretraining helps for many natural language understanding tasks. These examples of multi-task learning demonstrate the potential benefits of incorporating additional tasks in NLP models.

3 Models

To find an effective way to combine the models, we first fine-tune the individual models separately to find the optimal set of hyper-parameters for individual tasks. Moreover, we need SRL fine-tuned model as the input for the combined models. For ABSA, we adopt the model proposed by (Sun et al., 2019). We propose a new SRL end-to-end model, specifically designed for seamless integration with other tasks.

3.1 Semantic Role Labeling

Our goal is to train a universal encoder that effectively captures SRL information from a plain-text input. To accomplish this, we propose an end-to-end model with a single projection layer on the top of the ELECTRA encoder (or any other pre-trained language model). This way, all the information useful to predict role labels is encoded in the last hidden state of the encoder. Consequently, we can use this representation in other tasks. Although our end-to-end model exhibits lower performance than the commonly used BERT SRL model (Shi and Lin, 2019; Sido et al., 2021), we believe it is more suitable for this task.

In our end-to-end model, we first encode the whole sentence and then iterate over all possible word pairs (the first word is a potential predicate and the second is a potential argument). For each potential predicate-argument pair, we first concatenate the representations of predicate and argument and then classify the argument role. If the potential predicate is not a real predicate word or the potential argument is not an argument of the predicate, the role of the pair is set to *Other*. If a word is represented by multiple subword tokens, only the first token is classified. This is common practice in tagging tasks where the model learns to encode the semantics of a multi-token word into the first subword, then each word has a single token on the output for its classification.
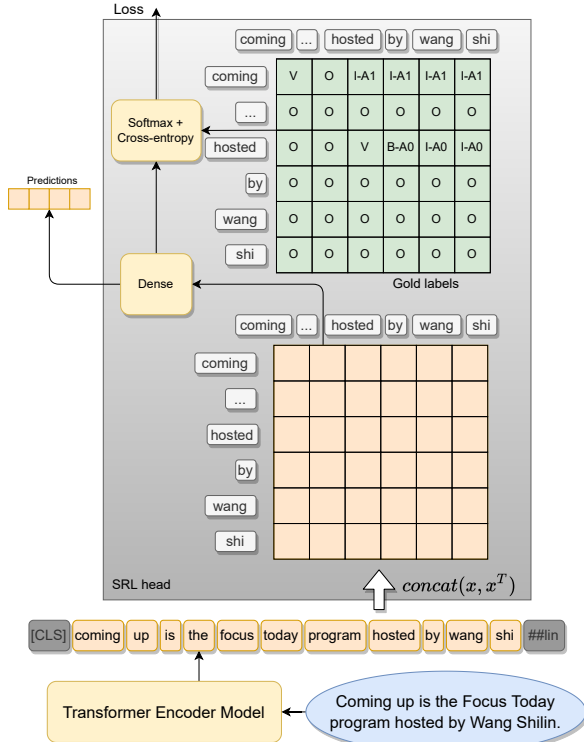
Figure 3: End-to-end SRL model architecture.

Our approach differs from that of Shi and Lin (2019) in terms of how the predicate-argument structure of the sentence is encoded within the transformer model. While Shi and Lin (2019) encodes each predicate separately and requires gold predicates on input, our model only requires plain text as input. In other words, our model requires only text as input, but the model proposed by Shi and Lin (2019) operates on pairs of text-predicate, producing representations solely for the input pair rather than the entire SRL output encompassing all predicates within the sentence. Figure 3 shows the schema of our end-to-end SRL model.

For our approach, it is necessary to have the same format of input (i.e., plain text) for both tasks that are combined. This is the reason why we need our end-to-end SRL model. For multitask learning, we need a general-purpose model, the same for both tasks. The task-specific models may yield better results on the SRL task, but they are specifically oriented only on the SRL task and makes their integration with ABSA or utilization in multitask learning challenging, if not impossible.

3.2 Aspect-Based Sentiment

As we mentioned in the introduction, we tackle the CE and CP subtasks of ABSA, as one classification task. We adopt the same approach as Sun et al. (2019), and we construct auxiliary sentences and convert the subtasks to a binary classification task.

We use the NLI-B approach from Sun et al. (2019) to build the auxiliary sentences. For each sentence, we build multiple auxiliary pseudo sentences that are generated for every combination of all polarity labels and aspect categories[3]. Each example has a binary label $l \in \{0, 1\}$; $l = 1$ if the auxiliary sentence corresponds to the original labels, $l = 0$ otherwise. We also add the artificial polarity class *none* that has assigned binary label $l = 1$ if there is no aspect category for a given sentence. The pseudo auxiliary sentence consists only of a polarity label and aspect category in a given language. For example, the auxiliary sentences for all aspects of the sentence "*The burger was excellent but the waitress was unpleasant*" are shown in Figure 4.

label		sentence	label		sentence
		food			service
1	⇒	positive – food	0	⇒	positive – service
0	⇒	negative – food	1	⇒	negative – service
0	⇒	neutral – food	0	⇒	neutral – service
0	⇒	conflict – food	0	⇒	conflict – service
0	⇒	none – food	0	⇒	none – service
		price			ambience
0	⇒	positive – price	0	⇒	positive – ambience
0	⇒	negative – price	0	⇒	negative – ambience
0	⇒	neutral – price	0	⇒	neutral – ambience
0	⇒	conflict – price	0	⇒	conflict – ambience
1	⇒	none – price	1	⇒	none – ambience
		general			
0	⇒	positive – general			
0	⇒	negative – general			
0	⇒	neutral – general			
0	⇒	conflict – general			
1	⇒	none – general			

Figure 4: Example of auxiliary sentences.

Each auxiliary sentence is combined with the original sentence and separated with [SEP] token and forms one training example, e.g., [CLS] *positive - food* [SEP] *the burger was excellent but the waitress was unpleasant* [SEP]. We fine-tune the pre-trained transformer model for the binary classification task on all generated training examples as Sun et al. (2019).

3.3 Combined Models

We propose several models designed to use SRL representation to enhance ABSA performance. The first type of model predicts aspect and sentiment

---

[3]For English we have four polarity labels plus artificial label *none* and five aspect categories, i.e. 25 possible auxiliary sentences. For Czech there is 20 possible sentences ($3 + 1$ polarity labels and five aspect categories).

using concatenated representations from both the SRL and ABSA encoders. The SRL encoder is pre-trained (pre-fine-tuned) on the SRL data, and its weights remain fixed during sentiment training. Since SRL is a token-level task, we need to reduce the sequential dimension before performing the concatenation step. To address this, we employ two approaches: simple average-over-time pooling (named *concat-avg*) and a convolution layer followed by max-over-time pooling (named *concat-conv*). Figure 5 shows the model architecture.
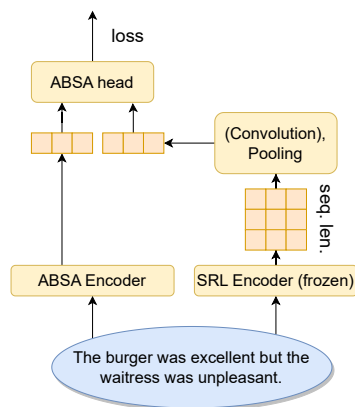

Figure 5: Concat model architecture.

The last model uses standard multi-task learning. We utilize a single Transformer encoder with two classification heads: one for the sentiment (standard head for sequence classification) and the other for SRL (the head architecture is presented in the previous section with the end-to-end SRL model). The model is trained using alternating batches, it means that we use different training data for both tasks, and we are not mixing them in a batch. In a single batch, we provide only ABSA or SRL data. See Figure 6 model's architecture.
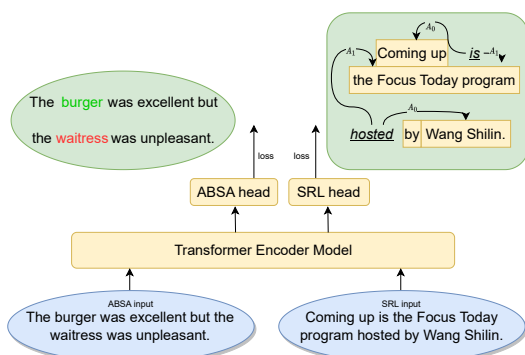

Figure 6: Multi-task model architecture.

4 Experiments

In our experiments, we aim to verify our idea that injected SRL information can improve the results of the ABSA task, particularly the CP subtask.

4.1 Datasets & Models Fine-Tuning

For Semantic Role Labeling, we use OntoNotes 5.0 dataset (Weischedel et al., 2013) for English and CoNLL 2009 (Hajic et al., 2009) for Czech. As metrics, we report the whole role F1 score for both datasets. Additionally, for English, we report CoNLL 2003 official score as a comparative metric as it is the standard metric used with OntoNotes.

For Aspect-Based Sentiment, we use the widely-used English dataset from Pontiki et al. (2014) that consists of 3,044 train and 800 test sentences from the restaurant domain. The English dataset contains four sentiment labels: *positive*, *negative*, *neutral*, and *conflict*. Further, we split[4] the original training part of 3,044 sentences into development (10%) and training parts (90%).

For Czech experiments, we employ the dataset from Hercig et al. (2016) with 2,149 sentences from the restaurant domain. Unlike in the English dataset, there are only three polarity labels: *positive*, *negative*, and *neutral*. Because the dataset has no official split, we divided[4] the data into training, development, and testing parts with the following ratio: 72% for training, 8% for the development evaluation, and 20% for testing. Both Czech and English datasets contain five aspect categories: *food*, *service*, *price*, *ambience*, and *general*.

For our experiments on English, we use the pre-trained *ELECTRA-small* model introduced by Clark et al. (2020), which has 14M parameters. For Czech, we employ the pre-trained monolingual model *Small-E-Czech* (Kocián et al., 2021) with the same size and architecture. Firstly, we train separate models for both tasks (ABSA and SRL) and select the optimal set of hyper-parameters on the development data. We then use the same hyper-parameters in combined models. For the details of hyper-parameters, see Appendix A.

4.2 Results & Discussion

We report the results of our end-to-end SRL model in Table 3. As we expected, our model performs worse than the model proposed by Shi and Lin (2019), but the results are reasonably high (con-

---

[4]For both English and Czech we provide a script to obtain the same split distribution.

Model	Category Extraction			Category Polarity	
	F1 Micro	Precision	Recall	Acc #3	Acc #2
baseline	$86.04^{\pm0.36}$	$86.48^{\pm0.97}$	$85.62^{\pm0.65}$	$75.58^{\pm0.55}$	$88.69^{\pm0.26}$
concat-conv	$\mathbf{86.58^{\pm0.54}}$	$\mathbf{86.90^{\pm0.51}}$	$\mathbf{86.28^{\pm0.94}}$	$\mathbf{79.20^{\pm0.48}}$	$\mathbf{90.26^{\pm0.58}}$
concat-avg	$86.34^{\pm0.57}$	$86.57^{\pm0.84}$	$86.12^{\pm1.08}$	$78.33^{\pm0.64}$	$90.06^{\pm0.79}$
multi-task	$85.62^{\pm0.63}$	$86.24^{\pm0.66}$	$85.01^{\pm0.66}$	$77.27^{\pm0.69}$	$89.00^{\pm0.63}$
baseline (Hercig et al., 2016)*	71.70	-	-	69.70	-
best (Hercig et al., 2016)*	80.00	-	-	75.20	-
CNN2 (Lenc and Hercig, 2016)	-	-	-	$69.00^{\pm2.00}$	-

Table 1: Czech results for the category extraction (CE) subtask as F1 Micro score, Precision and Recall. Results for the category polarity (CP) subtask as accuracy for three polarity labels (Acc #3) and binary polarity labels (Acc #2). Results marked with * symbol were obtained by 10-fold cross-validation.

Model	Category Extraction			Category Polarity		
	F1 Micro	Precision	Recall	Acc #4	Acc #3	Acc #2
baseline	$89.50^{\pm0.45}$	$90.95^{\pm0.70}$	$88.09^{\pm0.48}$	$83.03^{\pm0.43}$	$86.91^{\pm0.55}$	$92.74^{\pm0.53}$
concat-conv	$\mathbf{89.74^{\pm0.55}}$	$\mathbf{91.24^{\pm0.54}}$	$\mathbf{88.28^{\pm0.77}}$	$\mathbf{84.19^{\pm0.49}}$	$\mathbf{88.08^{\pm0.41}}$	$\mathbf{93.76^{\pm0.46}}$
concat-avg	$89.58^{\pm0.43}$	$91.15^{\pm0.60}$	$88.08^{\pm0.66}$	$84.13^{\pm0.51}$	$87.95^{\pm0.46}$	$93.49^{\pm0.44}$
multi-task	$89.36^{\pm0.15}$	$90.72^{\pm0.52}$	$88.05^{\pm0.44}$	$82.83^{\pm1.10}$	$87.05^{\pm1.21}$	$92.74^{\pm0.79}$
XRCE (Brun et al., 2014)	82.29	83.23	81.37	78.10	-	-
NRC (Kiritchenko et al., 2014)	88.58	91.04	86.24	82.90	-	-
BERT single (Sun et al., 2019)	90.89	92.78	89.07	83.70	86.90	93.30
NLI-B (Sun et al., 2019)	92.18	93.57	90.83	84.60	88.70	95.10
QACG-B (Wu and Ong, 2021)	92.64	$94.38^{\pm0.31}$	$\underline{90.97^{\pm0.28}}$	$\underline{86.80^{\pm0.80}}$	$90.10^{\pm0.30}$	$\underline{95.60^{\pm0.40}}$
BART generation (Liu et al., 2021)	$\underline{92.80}$	$\underline{95.18}$	90.54	-	$\underline{90.55^{\pm0.32}}$	-

Table 2: English results for the category extraction (CE) subtask as F1 Micro score, Precision and Recall. Results for category polarity (CP) subtask as accuracy for four polarity labels (Acc #4), three polarity labels (Acc #3) and binary polarity labels (Acc #2).

sidering that it does not have gold predicates on input).

Model	EN	EN-conll05	CS
(Shi and Lin, 2019)	88.89	85.20	83.09
end-to-end (ours)	84.54	81.51	79.74

Table 3: Comparison of results of the standard model and our end-to-end SRL model (reported in F1 scores, the official metrics, for the datasets used).

Results for our ABSA experiments in Czech and English are shown in Tables 1 and 2, respectively. The *baseline* refers to the model described in Section 3.2 without any injected SRL information. The SotA results are underlined and the best results for our experiments are bold. We include the results with the 95% confidence interval (experiments repeated 12 times). We use the F1 Micro and accuracy for the CE and CP subtasks, respectively. Based on the results presented in Tables 1 and 2, we can observe that our proposed models (*concat-conv* and *concat-avg*) with injected SRL information consistently enhance results for the CP subtask in

both languages. These improvements are statistically significant. The performance of the *concat-conv* and *concat-avg* models does not exhibit a significant difference. In the CE subtask, we achieve the same results as the *baseline* model. We think that the CE subtask is more distant from the SRL task than the CP subtask and therefore, the injection of the semantic information does not help. In other words, the semantic structure of the sentence may not play a crucial role in aspect detection (that can be viewed as multi-label text classification). On the other hand, for the CP subtask, the combined models can leverage the semantic structure of the sentence to their advantage.

For the Czech ABSA dataset we achieve new SotA results on both subtasks[5]. As we expected, we did not outperform the current SotA results for the English dataset, as our ELECTRA model has considerably fewer parameters than SotA models. For Czech, the *multi-task* model exhibited a marginal

---

[5]It is worth noting that although the test data we used differ from those used by Hercig et al. (2016) due to their 10-fold cross-validation, the performance difference is substantial enough to demonstrate the superiority of our approach.

improvement in the results and generally, the model was significantly inferior to our other models. We decided to use the smaller ELECTRA-based models because of their much smaller computation requirements. However, in future work, we plan comparison with larger models like BERT or RoBERTa to obtain the overall performance overview of our approach.

5 Conclusion

In this work, we introduce a novel end-to-end SRL model that we use to improve the aspect category polarity task. Our contribution lies in proposing several methods to integrate SRL and ABSA models, which ultimately lead to improved performance. The experimental results validate our initial assumption that leveraging semantic information extracted from an SRL model can significantly enhance the aspect category polarity task. Importantly, the approaches we propose are versatile and can be applied to combine Transformer-based models for other related tasks as well, extending the scope of their applicability.

Moreover, we believe that our approaches hold even greater potential in addressing other ABSA subtasks, namely term extraction and term polarity classification. These subtasks could benefit from the integration of SRL and ABSA models in a similar manner. Further, we would like to validate our approach on larger models, for example, BERT or RoBERTa.

Acknowledgements

References

Collin F. Baker, Charles J. Fillmore, and John B. Lowe. 1998. The Berkeley FrameNet project. In *Proceedings of the 17th International Conference on Computational Linguistics - Volume 1*, COLING '98, pages 86–90, Stroudsburg, PA, USA. Association for Computational Linguistics.

Caroline Brun, Diana Nicoleta Popa, and Claude Roux. 2014. XRCE: Hybrid classification for aspect-based sentiment analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 838–842, Dublin, Ireland. Association for Computational Linguistics.

Peng Chen, Zhongqian Sun, Lidong Bing, and Wei Yang. 2017. Recurrent attention network on memory for aspect sentiment analysis. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 452–461, Copenhagen, Denmark. Association for Computational Linguistics.

Kevin Clark, Minh-Thang Luong, Quoc V. Le, and Christopher D. Manning. 2020. ELECTRA: Pre-training text encoders as discriminators rather than generators. In *ICLR*.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. 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)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Feifan Fan, Yansong Feng, and Dongyan Zhao. 2018. Multi-grained attention network for aspect-level sentiment classification. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3433–3442, Brussels, Belgium. Association for Computational Linguistics.

Gayatree Ganu, Noémie Elhadad, and Amélie Marian. 2009. Beyond the stars: Improving rating predictions using review text content. In *WebDB*.

Daniel Gildea and Daniel Jurafsky. 2002. Automatic labeling of semantic roles. *Computational linguistics*, 28(3):245–288.

Jan Hajic, Massimiliano Ciaramita, Richard Johansson, Daisuke Kawahara, M Antònia Martí, Lluís Màrquez, Adam Meyers, Joakim Nivre, Sebastian Padó, Jan Štěpánek, et al. 2009. The conll-2009 shared task: Syntactic and semantic dependencies in multiple languages. In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning (CoNLL 2009): Shared Task*, pages 1–18.

Kazuma Hashimoto, Caiming Xiong, Yoshimasa Tsuruoka, and Richard Socher. 2016. A joint many-task model: Growing a neural network for multiple nlp tasks. *arXiv preprint arXiv:1611.01587*.

Luheng He, Kenton Lee, Omer Levy, and Luke Zettlemoyer. 2018. Jointly predicting predicates and arguments in neural semantic role labeling. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 364–369.

Luheng He, Kenton Lee, Mike Lewis, and Luke Zettlemoyer. 2017. Deep semantic role labeling: What works and what's next. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 473–483.

Tomáš Hercig, Tomáš Brychcín, Lukáš Svoboda, Michal Konkol, and Josef Steinberger. 2016. Unsupervised methods to improve aspect-based sentiment analysis in czech. *Computación y Sistemas*, 20(3):365–375.

Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation*, 9(8).

Minqing Hu and Bing Liu. 2004. Mining and summarizing customer reviews. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '04, page 168–177, New York, NY, USA. Association for Computing Machinery.

Diederik P. Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.

Svetlana Kiritchenko, Xiaodan Zhu, Colin Cherry, and Saif Mohammad. 2014. NRC-Canada-2014: Detecting aspects and sentiment in customer reviews. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 437–442, Dublin, Ireland. Association for Computational Linguistics.

Matěj Kocián, Jakub Náplava, Daniel Štancl, and Vladimír Kadlec. 2021. Siamese bert-based model for web search relevance ranking evaluated on a new czech dataset. *arXiv preprint arXiv:2112.01810*.

Jan Lehečka, Jan Švec, Pavel Ircing, and Luboš Šmídl. 2020. Bert-based sentiment analysis using distillation. In *Statistical Language and Speech Processing*, pages 58–70, Cham. Springer International Publishing.

Ladislav Lenc and Tomás Hercig. 2016. Neural networks for sentiment analysis in czech. In *ITAT*, pages 48–55.

Ming-Fan Li, Kaijie Zhou, Xuan Li, and Jianping Shen. 2021. Aspect-based sentiment classification with background information and syntactic auxiliary tasks. In *2021 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8.

Yuncong Li, Cunxiang Yin, Sheng-hua Zhong, and Xu Pan. 2020. Multi-instance multi-label learning networks for aspect-category sentiment analysis. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 3550–3560, Online. Association for Computational Linguistics.

Bin Liang, Hang Su, Lin Gui, Erik Cambria, and Ruifeng Xu. 2022. Aspect-based sentiment analysis via affective knowledge enhanced graph convolutional networks. *Knowledge-Based Systems*, 235:107643.

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

Jian Liu, Zhiyang Teng, Leyang Cui, Hanmeng Liu, and Yue Zhang. 2021. Solving aspect category sentiment analysis as a text generation task. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 4406–4416, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

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.

Dehong Ma, Sujian Li, Xiaodong Zhang, and Houfeng Wang. 2017. Interactive attention networks for aspect-level sentiment classification. *arXiv preprint arXiv:1709.00893*.

Alessandro Moschitti, Daniele Pighin, and Roberto Basili. 2008. Tree kernels for semantic role labeling. *Computational Linguistics*, 34(2):193–224.

Martha Palmer, Daniel Gildea, and Paul Kingsbury. 2005. The Proposition Bank: An annotated corpus of semantic roles. *Computational Linguistics*, 31(1):71–106.

Sean Papay, Roman Klinger, and Sebastian Padó. 2021. Constraining linear-chain crfs to regular languages. *arXiv preprint arXiv:2106.07306*.

Maria Pontiki, Dimitris Galanis, John Pavlopoulos, Harris Papageorgiou, Ion Androutsopoulos, and Suresh Manandhar. 2014. SemEval-2014 task 4: Aspect based sentiment analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 27–35, Dublin, Ireland. Association for Computational Linguistics.

Pavel Přibáň and Josef Steinberger. 2021. Are the multilingual models better? improving Czech sentiment with transformers. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021)*, pages 1138–1149, Held Online. INCOMA Ltd.

Pavel Přibáň and Josef Steinberger. 2022. Czech dataset for cross-lingual subjectivity classification. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 1381–1391, Marseille, France. European Language Resources Association.

Alexander Rietzler, Sebastian Stabinger, Paul Opitz, and Stefan Engl. 2020. Adapt or get left behind: Domain adaptation through BERT language model finetuning for aspect-target sentiment classification. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 4933–4941, Marseille, France. European Language Resources Association.

Peng Shi and Jimmy Lin. 2019. Simple bert models for relation extraction and semantic role labeling. *arXiv preprint arXiv:1904.05255*.

Jakub Sido, Ondřej Pražák, Pavel Přibáň, Jan Pašek, Michal Seják, and Miloslav Konopík. 2021. Czert – Czech BERT-like model for language representation. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021)*, pages 1326–1338, Held Online. INCOMA Ltd.

Josef Steinberger, Tomáš Brychcín, and Michal Konkol. 2014. Aspect-level sentiment analysis in Czech. In *Proceedings of the 5th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 24–30, Baltimore, Maryland. Association for Computational Linguistics.

Chi Sun, Luyao Huang, and Xipeng Qiu. 2019. Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence. 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)*, pages 380–385, Minneapolis, Minnesota. Association for Computational Linguistics.

Mihai Surdeanu, Richard Johansson, Adam Meyers, Lluís Màrquez, and Joakim Nivre. 2008. The conll 2008 shared task on joint parsing of syntactic and semantic dependencies. In *CoNLL 2008: Proceedings of the Twelfth Conference on Computational Natural Language Learning*, pages 159–177.

Zhixing Tan, Mingxuan Wang, Jun Xie, Yidong Chen, and Xiaodong Shi. 2018. Deep semantic role labeling with self-attention. In *Thirty-Second AAAI Conference on Artificial Intelligence*.

Duyu Tang, Bing Qin, Xiaocheng Feng, and Ting Liu. 2016. Effective LSTMs for target-dependent sentiment classification. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 3298–3307, Osaka, Japan. The COLING 2016 Organizing Committee.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.

Nan Wang, Jiwei Li, Yuxian Meng, Xiaofei Sun, and Jun He. 2021. An mrc framework for semantic role labeling. *arXiv preprint arXiv:2109.06660*.

Ralph Weischedel, Martha Palmer, Mitchell Marcus, Eduard Hovy, Sameer Pradhan, Lance Ramshaw, Nianwen Xue, Ann Taylor, Jeff Kaufman, Michelle Franchini, et al. 2013. Ontonotes release 5.0 ldc2013t19. *Linguistic Data Consortium, Philadelphia, PA*, 23.

Zhengxuan Wu and Desmond C Ong. 2021. Context-guided bert for targeted aspect-based sentiment analysis. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 14094–14102.

Hu Xu, Bing Liu, Lei Shu, and Philip Yu. 2019. BERT post-training for review reading comprehension and aspect-based sentiment analysis. 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)*, pages 2324–2335, Minneapolis, Minnesota. Association for Computational Linguistics.

Chen Zhang, Qiuchi Li, and Dawei Song. 2019. Aspect-based sentiment classification with aspect-specific graph convolutional networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 4568–4578, Hong Kong, China. Association for Computational Linguistics.

Yu Zhang, Qingrong Xia, Shilin Zhou, Yong Jiang, Zhenghua Li, Guohong Fu, and Min Zhang. 2021. Semantic role labeling as dependency parsing: Exploring latent tree structures inside arguments. *arXiv preprint arXiv:2110.06865*.

Zhuosheng Zhang, Yuwei Wu, Hai Zhao, Zuchao Li, Shuailiang Zhang, Xi Zhou, and Xiang Zhou. 2020. Semantics-aware bert for language understanding. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 9628–9635.

A   Training Hyper-Parameters

We use the Adam (Kingma and Ba, 2015) optimizer with default parameters ($\beta_1 = 0.9, \beta_2 = 0.999$) and the cross-entropy loss function for all our experiments. The initial learning rate is set to 2e-5 with linear decay to zero. We fine-tune all models with batch size 32 and maximum sequence length 256. All data fit into this length. The models are trained for 120 epochs in Czech and 40 in English. The epochs are measured in ABSA data. The multi-task model is trained on the same amount of SRL data additionally (because we use alternating batches).

B   Semantic Parse Tree Example

As mentioned in the introduction section, we assume that leveraging SRL information can prove advantageous in the aspect category polarity (CP) task. To illustrate this point, consider the annotation depicted in Figure 8, where we can observe the SRL relation extracted (see Figure 7) between the words *forgotten* and *food*. The information about this relation can help to understand the model that these words are related and help the model to predict the negative polarity of the food aspect category.
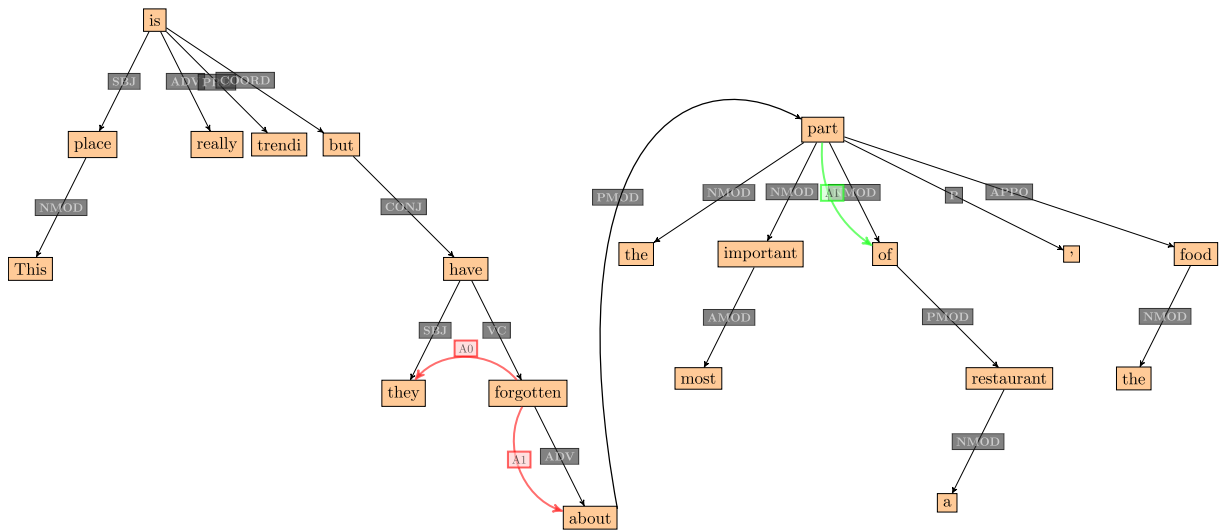
Figure 7: Example of syntactic and semantic parse tree of the following sentence "*This place is really trendi but they have forgotten about the most important part of a restaurant, the food*".

"*This place is really trendy but they have forgotten about the
    most important part of a restaurant, the food.* "
    CE ⇒ food, ambience
    CP ⇒ food:*negative*, ambience:*positive*

Figure 8: Example of CE and CP annotations.