

Aspect Based Sentiment Analysis using French Pre-Trained Models

Abderrahman Essebbar, Bamba Kane, Ophélie Guinaudeau, Valeria Chiesa, Ilhem Quénel
and Stéphane Chau

Research and Innovation Direction, ALTRAN Sophia-Antipolis, France

Keywords: Natural Language Processing, Aspect Based Sentiment Analysis, Sentiment Analysis, Pre-Trained Models, Sentence Pair Classification, Attention Encoder Network, Aspect Sentiment Classification, SemEval.

Abstract: Aspect Based Sentiment Analysis (ABSA) is a fine-grained task compared to Sentiment Analysis (SA). It aims to detect each aspect evoked in a text and the sentiment associated to each of them. For English language, many works using Pre-Trained Models (PTM) exists and many annotated open datasets are also available. For French Language, many works exists in SA and few ones for ABSA. We focus on aspect target sentiment analysis and we propose an ABSA using French PTM like multilingual BERT (mBERT), CamemBERT and FlauBERT. Three different fine-tuning methods: Fully-Connected, Sentences Pair Classification and Attention Encoder Network, are considered. Using the SemEval2016 French reviews datasets for ABSA, our fine-tuning models outperforms the state-of-the-art French ABSA methods and is robust for the Out-Of-Domain dataset.

1 INTRODUCTION

Aspect Based Sentiment Analysis (ABSA) helps businesses to become more and more customer-centric. It is used as a tool to deep understand customers by analyzing their feedback and expectations. It allows to point out the satisfactory aspects and those that need to be improved based on the customer experience (Pang and Lee, 2008).

Over the past few years, ABSA has been developed for several applications : movie reviews, customer reviews on electronic products (e.g. cameras, computers), services, restaurants etc.

ABSA is part of Natural Language Processing (NLP) and it is well known to provide more information about the context than a simple sentiment analysis (SA). There has been a large amount of work in SA over the last decade and it continues to rapidly grow in new directions (Lin and Luo, 2020). A major issue in this field is that a customer review can express sentiments towards various aspects of a product or service. For example, a restaurant review can talk positively about the food, and also talk negatively about the price of the menu. Thus, SA is not enough.

ABSA aims to split the text into *Aspects* (attributes or components of a product or service) and then give to each aspect a *Sentiment* level: positive, negative or neutral.

An example of restaurant review showing the two

aspects category and target (or term) (Apidianaki et al., 2016): [*Pourtant les plats sont bons et la deco est sympa* (However, the food is good and the decoration is nice)]

[category=FOOD#QUALITY,target:*plats*,polarity : positive]

[category=AMBIENCE#GENERAL,target:*deco*, polarity : positive]

There are different ABSA tasks:

- The Opinion Target Extraction which aims at extracting the target (i.e the words reflecting the aspect);
- The Aspect-Category Detection which aims at detecting the different types of aspects that are evoked in a text.;
- The Aspect Sentiment Classification (ASC) for which the objective is to associate a sentiment polarity (positive, negative or neutral) to each identified aspect.

In this paper, we focus on **ASC**.

For this kind of tasks, Pre-Trained Models (PTM) appear to be very promising solutions but until now they have not been used for ABSA in French Language. The main contributions of this research work are:

- Proposition of the first work on ABSA using French language PTM whereas many ABSA

works are done for English language ;

- Adaptation of three fine-tuning methods ;
- Comparison with the state-of-the-art methods using the first french SemEval2016 dataset for Restaurant and Museum (Apidianaki et al., 2016)

The paper is structured as follows: a review of ABSA state-of-the-art methods is presented in section 2, while in section 3 the french PTM are introduced. In section 4, a benchmark study of the results is exposed, together with the dataset used for SemEval2016 challenge in French language.

2 STATE-OF-THE-ART ABSA METHODS

2.1 Conventional Methods

In this section, we present the state-of-the-art for ASC using conventional classifiers like Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), Conditional Random Field (CRF):

- (Brun et al., 2016) proposed a method composed by two steps : 1) at word level, a CRF is trained to classify terms in aspect categories; 2) at sentence level, aspect categories are associated to a sentence according with a probability . This method is the winner for French language SemEval2016 Challenge (slot 1 and slot 3) (Pontiki et al., 2016).
- (Kumar et al., 2016) used information extracted from dependency graphs learned on different domains and languages of SemEval2016 and showed very efficient results on different languages including French language .
- (Macháček, 2016) focused only on aspect categories and modeled the task as a multi-label classification with binary relevance transformation, where labels correspond to the aspects.
- (Ruder et al., 2016) proposed a method using multiple CNN filters for sentiment and aspect detection.
- (Tang et al., 2016b) applied a deep Memory Network (**MemNet**) which uses multiple attention to compute the importance of each context word .
- Target-Dependent LSTM (**TD-LSTM**): (Tang et al., 2016a) used LSTM networks to model both the left context and the right context with the respect to given target. Then the left and right target-dependent representations are concatenated to predict the sentiment polarity of the target.

- Target-Connection LSTM (**TC-LSTM**): method uses a LSTM for which semantic relatedness of target with its context words are incorporated (Tang et al., 2016a). A target vector is calculated by averaging the vectors of the words that compose the target.
- (Kooli and Pigeul, 2018) propose the CNN-LSTM-CRF model for aspects detection and the MEMNet model for detecting the sentiment associated with the aspects. these separate methods are applied on French SemEval2016 data.

The previous standard LSTM based models cannot detect which is the important part for ASC. When classifying the polarity of one sentence given the aspect, the aspect information is important. We may get opposite polarities if different aspects are considered. The use of target information only is not sufficient. The application of attention mechanism can extract the association of important words for an aspect (Wang et al., 2016), and can capture the key part of sentences in response to a given aspect.

- Attention-based LSTM with Aspect Embedding (**ATAE-LSTM**): model appends the target embeddings with each word embeddings and uses BiLSTM with attention to get the aspect and the associated sentiment (Wang et al., 2016).
- Interactive Attention Network (**IAN**): model aims to learn the representations of the target and context with LSTM and attentions interactively, which generates the representations for targets and contexts with respect to each other (Ma et al., 2017).
- Attentional Encoder Network (**AEN**): model proposed by (Song et al., 2019) avoids recurrence and employs attention based encoders for modeling context and aspect.

2.2 PTM Methods

The evolution of word representation used in NLP started with non-neural methods, neural word embedding techniques, context word embedding methods and actually the trend is large pre-trained language models like Bidirectional Encoder Representations from Transformers (BERT) and others (Qiu et al., 2020). The PTM provide a context to words that have previously been learning the occurrence and representations of words from unannotated training data.

BERT is a pre-trained English language model that is designed to consider the context of a word from both left and right side simultaneously (Devlin et al., 2018). BERT is not based on LSTM to get the word

Table 1: PTM parameters comparison (from (Le et al., 2019)).

	BERT _{BASE}	RoBERTa _{BASE}	CamemBERT	FlauBERT _{BASE} /FlauBERT _{LARGE}
Language	English	English	French	French
Training data	13 GB	160 GB	138 GB [†]	71 GB [‡]
Pre-training objectives	NSP and MLM	MLM	MLM	MLM
Total parameters	110 M	125 M	110 M	138 M/ 373 M
Tokenizer	WordPiece 30K	BPE 50K	SentencePiece 32K	BPE 50K
Masking strategy	Static + Sub-word masking	Dynamic + Sub-word masking	Dynamic + Whole-word masking	Dynamic + Sub-word masking

[†], [‡]: 282 GB, 270 GB before filtering/cleaning.

context features, but instead uses transformers proposed by (Vaswani et al., 2017), which are attention-based mechanisms that are not based on recurrence.

BERT embedding generates word vectors of sequence in order to facilitate the training and Fine-Tuning for a specific task without having to make a major change in the model and parameters. This concept is simple and improves results for many NLP tasks such as SA and Question Answering (Q&A) systems, Part-Of-Speech (POS), Named Entity Recognition (NER) and NLI (Natural Language Inference).

For SA, BERT outperforms previous state-of-the-art models by simply fine-tuning on Stanford Sentiment Treebank and Internet Movie Database binary classification, which are widely used dataset for SA¹.

For ABSA, BERT with fine-tuned methods using Fully-Connected (FC) called also BASE, AEN or Sentence Pair Classification (SPC) shows improvements compared to conventional methods (Song et al., 2019) (Gao et al., 2019)

3 FRENCH PTM FOR ABSA

3.1 Existing French PTM

Based on the impact of PTM on NLP tasks in English, some work has recently released PTM for other languages and mainly in French language :

- **mBERT**: is a multilingual BERT with many languages including french language (Pires et al., 2019).
- **CamemBERT**: is the first monolingual PTM for French Language which is based on RoBERTa model (Facebook) (Martin et al., 2020).
- **FlauBERT**: is the monolingual PTM which is based on BERT model (Google)(Le et al., 2019). It was trained on almost twice as fewer text data than CamemBERT model.

¹http://nlpprogress.com/english/sentiment_analysis.html

The different PTM parameters are described in Table 1.

(Blard, 2020) proposed CamemBERT model for SA using French movies reviews scraped from the website *www.allocine.fr*. SA accuracy is improved for about three points compared to state-of-the-art methods. The fine tuning using CamemBERT also reduces the training dataset size.

(Le et al., 2019) applied FlauBERT model for SA on books, DVD and music French reviews. Their results show good performance even with small dataset.

Like BERT, the monolingual French PTM CamemBERT and FlauBERT improved the state-of-the-art performances for different NLP tasks (POS, NER, NLI, SA, Q&A). There results are also better compare to multilingual mBERT.

3.2 Our Proposition: Fine-tuning PTM for ABSA

Following, we propose three ABSA fine-tuning methods using French language PTM (mBERT, CamemBERT and FlauBERT):

- **PTM-FC** (Pre-Trained Model - Fully-Connected): Figure 1 shows the corresponding architecture. This fine-tuned method does not take into account the target information.
- **PTM-SPC** (Pre-Trained Model - Sentence Pair Classification): is used for many tasks (Devlin et al., 2018) and deals with determining the semantic relations between two sentences by taking two texts (sentence and targets) as input and outputs a label representing the type of relation between them. Figure 2 shows the corresponding architecture model.
- **PTM-AEN** (Pre-Trained Model - Attention Encoded Network): uses a PTM with AEN fine-tuning part. This method was proposed by Song et al. with English reviews (Song et al., 2019). The corresponding architecture is presented on Figure 3.

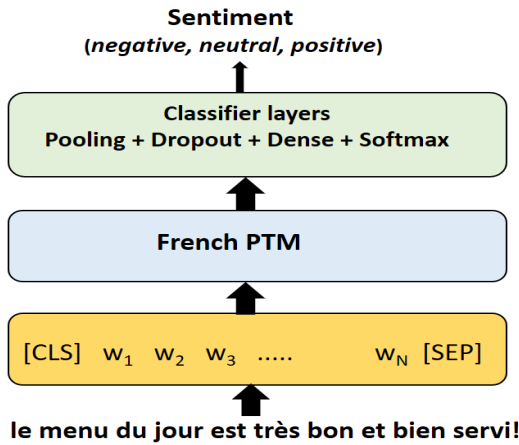


Figure 1: PTM-FC model architecture.

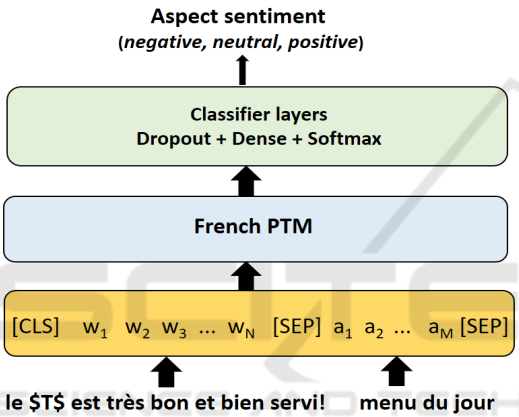


Figure 2: PTM-SPC model architecture.

4 EXPERIMENT

In this section, we evaluate the performances of different French PTM: CamemBERT, FlauBERT and mBERT. Moreover, for each PTM, we compare different fine-tuning methods: FC, SPC and AEN. We evaluate all the results and compare them with conventional methods (TD-LSTM, TC-LSTM, ATAE-LSTM, MEMNet and IAN).

4.1 SemEval2016 Dataset

For SA, there are many open dataset in French Language. ABSA dataset are very expensive and time-consuming to annotated manually. We apply our fine-tuned models on SemEval2016 french datasets about restaurants and museums (Apidianaki et al., 2016) (Pontiki et al., 2016).

The dataset description is given in table 2. The **Restaurant dataset** consists of 333 French reviews

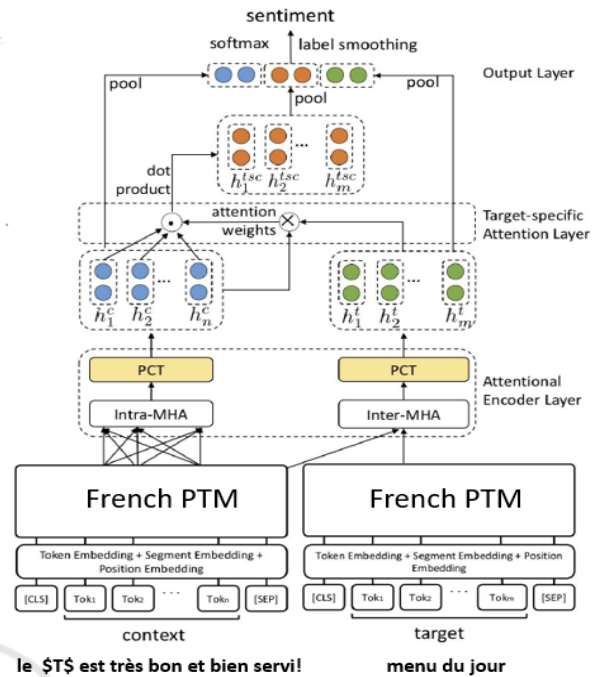


Figure 3: PTM-AEN model architecture (from (Song et al., 2019)).

annotated with targets, aspect categories and polarities (negative, neutral, positive) for training and 120 reviews for testing. There are 1660 annotated sentences for training and 696 for testing. The **Museum dataset** consists of 162 french reviews annotated with 668 sentences for testing. Table 3 describe the different aspect and category for both datasets.

Table 2: French data description (ABSA SemEval2016).

TRAINING	Texts	Sentences	Aspects
Restaurant	335	1669	1797
TESTING	Texts	Sentences	Aspects
Restaurant	120	696	708
Museum	162	686	582

Table 3: Aspect and category for Restaurant and Museum datasets.

Aspect	Category	Aspect	Category
RESTAURANT	GENERAL	MUSEUM	GENERAL
	PRICES		PRICES
	MISCELLANEOUS		MISCELLANEOUS
FOOD	QUALITY		COMFORT
	PRICES		ACTIVITIES
	STYLES & OPTIONS		ARCHITECTURE
DRINK	QUALITY	FACILITIES	GENERAL
	PRICES		PRICES
	STYLES & OPTIONS	COLLECTIONS	INTEREST
AMBIENCE	GENERAL		SET_UP
SERVICE	GENERAL	SERVICE	GENERAL
LOCATION	GENERAL	LOCATION	GENERAL
		TOUR_GUIDING	GENERAL

From each dataset, we construct a new dataset (Table 4) with different inputs: the text review with T , denot-

ing the aspect target and their corresponding polarity : positive, negative and neutral. Table 4 describes an example.

Table 4: Example of input data.

Review	Aspect Target	Polarity
une \$T\$ correcte	carte	neutral
le \$T\$ est très bon!	menu du jour	positive
\$T\$ minable !	accueil	negative

4.2 Training

For training and evaluation, the used hyperparameters are given in Table 5.

Conventional methods with word embedding representation uses a French Wikipedia2Vec with 300-dimensional embeddings trained on words from Wikipedia in French language ².

The French PTM with context word representation use the BASE (uncased) version for all our experiments.

The evaluation utilizes two measures: the accuracy (Acc) and the macro-F1 score (F1).

Table 5: Used hyperparameters.

Parameter	PTM	Conventional
Dropout rate	0.1	0.1
Batch size	16	16
Max seq length	80	80
Epochs number	5	5
Optimizer	Adam	Adam
Learning rate	$2.e^{-5}$	$5.e^{-4}$

4.3 Results for ABSA with Fine-tuned French PTM

The comparison of our models with state-of-the-art ASC methods on SemEval2016 french restaurants reviews is presented in Table 6.

The French PTM (mBERT, CamemBERT and FlauBERT) perform better compared to conventional methods which demonstrates the effectiveness of fine-tuning PTM on the ASC task.

We tested also the proposed models in a previously unseen Out-Of-Domain (OOD) Museum French reviews dataset. Table 3 shows that for both Museum and Restaurants reviews, testing data has only in common three aspects (RESTAURANT, LOCATION, SERVICE) and three categories (GENERAL, PRICES, MISCELLANEOUS). The

²<https://wikipedia2vec.github.io/wikipedia2vec/pretrained/>

Table 6: ASC for French SemEval2016 (Restaurants). CamemBERT-SCP offers the best performance for ABSA (all the PTM-FC performs only SA).

Method	Acc (%)	F1 (%)
Conventional Methods		
TC-LSTM	68.38	49.53
ATAE-LSTM	68.66	48.98
TD-LSTM	69.92	54.49
MEMNet	71.73	54.42
IAN	71.87	52.48
(Kooli and Pigeul, 2018)	74.23	-
(Brun et al., 2016)	78.82	-
Fine-tuned PTM		
mBERT-FC	79.06	67.16
mBERT-SCP	81.06	67.78
mBERT-AEN	80.22	66.70
CamemBERT-FC	83.5	70.88
CamemBERT-SCP	84.12	72.49
CamemBERT-AEN	83.98	71.54
FlauBERT-FC	84.68	73.88
FlauBERT-SCP	83.30	68.90
FlauBERT-AEN	84.06	71.70

Table 7: ASC for French SemEval2016 (Museums). CamemBERT-SCP and FlauBERT-SCP offers the best performance for ABSA (all the PTM-FC performs only SA).

Method	Acc (%)	F1 (%)
Conventional Methods		
TC-LSTM	63.06	40.33
ATAE-LSTM	64.26	42.26
TD-LSTM	65.64	44.90
MEMNet	66.49	43.29
IAN	68.38	43.87
(Kooli and Pigeul, 2018)	66.70	-
Fine-tuned PTM		
mBERT-FC	76.98	56.88
mBERT-SCP	76.80	53.76
mBERT-AEN	76.26	52.97
CamemBERT-FC	79.97	55.57
CamemBERT-SCP	81.62	59.09
CamemBERT-AEN	81.27	59.53
FlauBERT-FC	81.16	63.55
FlauBERT-SCP	80.24	62.80
FlauBERT-AEN	81.29	58.55

detailed results for museum dataset are shown in table 7. In this context also, the fine-tuned PTM methods results outperform the conventional methods.

As for others NLP tasks, the monolingual PTM CamemBERT and FlauBERT improve generally ABSA results around 3 points compared to mul-

lingual mBERT (except for the one case in table 7 for which mBERT-FC is actually better than CamemBERT-FC in term of F1-Score).

For ABSA task, FlauBERT and CamemBERT models shows comparable results as for others NLP tasks (Le et al., 2019). Flaubert shows then great interest since, as observed in Table 1, FlauBERT model was trained with half text data size compared to CamemBERT model.

The AEN model has the more complex structure and is not bringing significant improvements compare to SPC model.

For English language, the PTM BERT has shown great improvement on NLP tasks compared to the state-of-the-art. For ABSA, the improvement is smaller.

Data augmentation has shown the improvement of the performances with additional training on the review text. Many authors use data augmentation like BERT-PT (BERT Post-Training) with review reading comprehension (Xu et al., 2019) leading to improved performances. Adding Auxiliary Question BERT-AQ (Sun et al., 2019)) could also be a complementary way to improve aspect detection.

5 CONCLUSION

French PTM shows improvements of text representation in many NLP tasks including sentiment analysis at sentence-level. We propose the use of PTM for ABSA.

In this paper, we give an overview of the state-of-the-art methods for ABSA on French language. We propose and implement three fine-tuning methods (FC, SPC and AEN) using the French SemEval2016 data.

Experimental results showed that these methods outperforms conventional models with a word embedding representation. These results also indicate the higher performance of monolingual French models (FlauBERT and CamemBERT) compared to multilingual model (mBERT).

The PTM-SPC model shows generally great performances and is less complex compare to the PTM-AEN model. Our fine-tuned French PTM for ABSA are also robust for OOD Museum dataset.

For future work, we plan to explore other fine-tuned models and also to use data augmentation techniques with French PTM in order to improve the performances of our models.

REFERENCES

- Apidianaki, M., Tannier, X., and Richart, C. (2016). Datasets for aspect-based sentiment analysis in French. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 1122–1126, Portorož, Slovenia. European Language Resources Association (ELRA).
- Blard, T. (2020). French sentiment analysis with BERT. In <https://github.com/TheophileBlard/french-sentiment-analysis-with-bert>.
- Brun, C., Perez, J., and Roux, C. (2016). XRCE at SemEval-2016 task 5: Feedbacked ensemble modeling on syntactico-semantic knowledge for aspect based sentiment analysis. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 277–281, San Diego, California. Association for Computational Linguistics.
- Devlin, J., Chang, M., Lee, K., and Toutanova, K. (2018). BERT: pre-training of deep bidirectional transformers for language understanding. *CoRR*, abs/1810.04805.
- Gao, Z., Feng, A., Song, X., and Wu, X. (2019). Target-dependent sentiment classification with bert. In *IEEEAccess* 8864964.
- Kooli, N. and Pigeul, E. (2018). Analyse de sentiments à base d'aspects par combinaison de réseaux profonds : application à des avis en français. In *Actes de la Conférence TALN, Rennes*.
- Kumar, A., Kohail, S., A.Kumar, Ekbal, A., and Biemann, C. (2016). Beyond sentiment lexicon: Combining domain dependency and distributional semantics features for aspect based sentiment analysis. In *ACLweb* www.aclweb.org/anthology/S16-1174/.
- Le, H., Vial, L., Frej, J., Segonne, V., Coavoux, M., Lecouteux, B., Allauzen, A., Crabbé, B., Besacier, L., and Schwab, D. (2019). Flaubert: Unsupervised language model pre-training for french. *CoRR*, abs/1912.05372.
- Lin, P. and Luo, X. (2020). A survey of sentiment analysis based on machine learning. In Zhu, X., Zhang, M., Hong, Y., and He, R., editors, *Natural Language Processing and Chinese Computing - 9th CCF International Conference, NLPCC 2020, Zhengzhou, China, October 14-18, 2020, Proceedings, Part I*.
- Ma, D., Li, S., Zhang, X., and Wang, H. (2017). Interactive attention networks for aspect-level sentiment classification. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*, pages 4068–4074.
- Macháček, J. (2016). BUTknot at SemEval-2016 task 5: Supervised machine learning with term substitution approach in aspect category detection. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 301–305, San Diego, California. Association for Computational Linguistics.
- Martin, L., Muller, B., Ortiz Suárez, P. J., Dupont, Y., Romary, L., de la Clergerie, É., Seddah, D., and Sagot, B. (2020). CamemBERT: a tasty French language model. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7203–7219, Online. Association for Computational Linguistics.

- Pang, B. and Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval*, 2(1–2):1–135.
- Pires, T., Schlinger, E., and Garrette, D. (2019). How multilingual is multilingual BERT? In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4996–5001, Florence, Italy. Association for Computational Linguistics.
- Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., AL-Smadi, M., Al-Ayyoub, M., Zhao, Y., Qin, B., De Clercq, O., Hoste, V., Apidianaki, M., Tannier, X., Loukachevitch, N., Kotelnikov, E., Bel, N., Jiménez-Zafra, S. M., and Eryiğit, G. (2016). SemEval-2016 task 5: Aspect based sentiment analysis. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 19–30, San Diego, California. Association for Computational Linguistics.
- Qiu, X., Sun, T., Xu, Y., Shao, Y., Dai, N., and Huang, X. (2020). Pre-trained models for natural language processing: A survey. In *ArXiv preprint <https://arxiv.org/abs/2003.08271>*.
- Ruder, S., Ghaffari, P., and Breslin, J. G. (2016). INSIGHT-1 at SemEval-2016 task 5: Deep learning for multilingual aspect-based sentiment analysis. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 330–336, San Diego, California. Association for Computational Linguistics.
- Song, Y., Wang, J., Jiang, T., Liu, Z., and Rao, Y. (2019). Attentional encoder network for targeted sentiment classification. *CoRR*, abs/1902.09314.
- Sun, C., Huang, L., and Qiu, X. (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.
- Tang, D., Qin, B., Feng, X., and Liu, T. (2016a). 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.
- Tang, D., Qin, B., and Liu, T. (2016b). Aspect level sentiment classification with deep memory network. *CoRR*, abs/1605.08900.
- Vaswani, A., N. Shazeer, N. P., Uszkoreit, J., Jones, L., Gomez, A., L.Kaiser, and Polosukhin, I. (2017). Attention is all you need. In *ArXiv preprint 1706.03762*.
- Wang, Y., Huang, M., Zhu, X., and Zhao, L. (2016). Attention-based LSTM for aspect-level sentiment classification. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 606–615, Austin, Texas. Association for Computational Linguistics.
- Xu, H., Liu, B., Shu, L., and Yu, P. (2019). Bert post-training for review reading comprehension and aspect-based sentiment analysis. In *Proceedings of the 2019 Conference of the NAACL, pages 2324–2335, June 2019, Minneapolis, USA*.