CNN-LSTM-CRF for Aspect-Based Sentiment Analysis: A Joint Method Applied to French Reviews

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

Research and Innovation Direction, ALTRAN Sophia-Antipolis, France

bamba.kane, ali.jrad, abderrahman.essebbar, ophelie.guinaudeau, valeria.chiesa, ilhem.quenel,

Keywords: Natural Language Processing (NLP), Aspect-Based Sentiment Analysis (ABSA), Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), Conditional Random Field (CRF), SemEval.

Abstract: Aspect Based Sentiment Analysis (ABSA) aims to detect the different aspects addressed in a text and the sentiment associated to each of them. There exists a lot of work on this topic for the English language, but only few models are adapted for French ABSA. In this paper, we propose a new model for ABSA, named *CLC*, which combines CNN (Convolutional Neural Network), Bidirectional LSTM (Long Short-Term Memory) and CRF (Conditional Random Field). We demonstrate herein its great performance on the SemEval2016 French dataset. We prove that our CLC model outperforms the state-of-the-art models for French ABSA. We also prove that *CLC* is well adapted for other languages such as English. One main strength of *CLC* is its ability to detect the aspects and the associated sentiments in a joint manner, unlike the state-of-the-art models which detect them separately.

1 INTRODUCTION

Over the last few years a lot of work has been done in the field of sentiment analysis (Lin and Luo, 2020). Most of the work focuses on identifying the sentiment that is expressed in a text, in a global way by giving a positive, a negative or a neutral appreciation to the whole text. However, such a task is sometimes too coarse. For example, given the following review on a restaurant, *"the meat was delicious but the price was exaggerated"*, it is quite difficult to highlight an overall sentiment. Aspect Based Sentiment Analysis (ABSA) is a sub-domain of Natural Langage Processing (NLP) (Thet et al., 2010) which takes up this challenge and provides a much finer analysis by identifying the sentiment associated to each aspect evoked in the text.

In the previous example, ABSA objective is to detect both the two categories of aspects addressed in the text (*food_quality* and *food_price*), and the sentiment associated to each aspect (positive for *food_quality* and negative for *food_price*). A word (or a group of words) denoting an aspect is called target. In this review, *"the meat was delicious but the price was exaggerated"*, the target for the aspect *food_quality* is *meat*, and the one for *food_price* is *price*.

ABSA can be split into three tasks (Pontiki et al., 2016):

- Opinion Target Extraction (OTE): aims to detect the words that constitute the targets.
- Aspect Category Detection (ACD): focuses on the detection of the different types of aspects that are evoked in a text.
- Aspect Sentiment Polarity (ASP): assigns a sentiment polarity (positive, negative or neutral) for each identified aspect.

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

There is an increased interest in ABSA especially with the dedicated challenges of SemEval (Pontiki et al., 2016) that offers a framework to define and evaluate models on different domains. It was initiated for English language and it is now open to several other languages such as French or Spanish (Pontiki et al., 2016).

In order to detect the aspects and the related sentiments, many methods have been proposed. In (Kiritchenko et al., 2014) the authors use Support Vector Machine (SVM) for this purpose, while (Hamdan et al., 2015) use Conditional Random Field (CRF). Convolutional Neural Networks (CNN) have also

Kane, B., Jrad, A., Essebbar, A., Guinaudeau, O., Chiesa, V., Quénel, I. and Chau, S.
CNN-LSTM-CRF for Aspect-Based Sentiment Analysis: A Joint Method Applied to French Reviews.
DOI: 10.5220/0010382604980505
In *Proceedings of the 13th International Conference on Agents and Artificial Intelligence (ICAART 2021) - Volume 1*, pages 498-505
ISBN: 978-989-758-484-8

been widely employed for ABSA (Ruder et al., 2016; Wu et al., 2016).

However, most of these works focus on English language and there is only a few work in the literature for the French language. The lack of linguistic tools or resources is probably the explanation. Since the French language has it own characteristics such as gender for nouns, liaisons, or prepositions which are very different from English, the existing resources are not suitable.

In this article, we propose a new model named *CLC* which combines a CNN to encode word-level information into its word-level representation and a bidirectional Long Short Term Memory (BiLSTM) to model context information of each word. Finally, we use a sequential CRF to get the best sequence of aspects and the related sentiments. Our proposed approach leads to three main contributions:

- a new architecture for ABSA, named *CLC*, allowing to detect the aspects and associated sentiment in a joint manner;
- better performance than state-of-the-art methods for ABSA on the French language;
- great performance on English SemEval2016 dataset.

The paper is organized as follows: in section 2, we describe the state-of-the-art methods for ABSA. In section 3, the new model *CLC* is detailed. In section 4, the experimentation details are given. We show and analyze the performance on the SemEval2016 French dataset about restaurants and we provide a comparison with the state-of-the-art methods.

2 STATE OF THE ART

2.1 Aspect Based Sentiment Analysis (ABSA)

In this section, we present the state-of-the-art methods for ABSA and then we detail the methods that are relevant for French language.

2.1.1 Traditional Methods for ABSA

- **NRC-Canada:** (Kiritchenko et al., 2014) deployed a traditional Support Vector Machine (SVM)-based model and added some extensive feature engineering like Part of Speech (POS) tags, various types of n-grams, and lexicon features.
- **Rec-NN** (Recursive Neural Networks)**:** (Dong et al., 2014) used, in a first step, rules in order to get the word related to the aspect at the top of a dependency tree. Then, the representation of the sentence towards the aspect was learned by a semantic composition using Recursive Neural Network.
- **DDG** (Domain Dependency Graph)**:** (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 (among which the French language).

The two followings methods participate to the SemEval2016 challenge for French restaurant reviews:

- **XRCE:** it was the challenge winner for French language (Brun et al., 2016; Brun and Nikoulina, 2018). Their method was composed by two steps : 1) Classification process is performed using the output of a Conditional Random Field (CRF) which has been specialized at word level on the available training data (labeled aspects and sentiments) in order to classify terms into one or more aspect categories. 2) At sentence level, the classification models associate aspect categories of sentences with probabilities. The aspect categories are then assigned using a threshold over the assigned probabilities
- **BUTknot** (Machacek, 2016)**:** this method got the second rank for the SemEval2016 challenge in French. The aspect detection is based on supervised machine learning using bi-grams bag-of-words model for multi-languages. The performance is enhanced by a term substitution technique. The system has reached a very good performance in comparison with other submitted systems for the challenge.

2.1.2 Deep Learning Methods for ABSA

Many deep learning architectures are used for ABSA (Young et al., 2018):

- **CNN** (Ruder et al., 2016)**:** proposed a method using multiple filters and obtained competitive results on both polarity and aspect detection tasks. (Poria et al., 2016) designed a seven-layer CNN architecture and made use of both POS tagging and word embeddings as features.
- **MEM-Net** (Memory Network)**:** uses the inter-aspect relations modeling (Tang et al., 2016b). This method was used on French SemEval2016 datasets.

LSTM for ABSA. The LSTM model has achieved many great successes in various NLP tasks and particularly for ABSA. Following, we present different models for ABSA using LSTM (Tang et al., 2016a):

- **TD-LSTM** (Target-Dependent LSTM)**:** it defines the use of LSTM by considering the target. The BiLSTM network is used to model both the left and the right context of a target. Then the left and right target-dependent representations are concatenated for predicting the sentiment polarity of the target.
- **TC-LSTM** (Target-Connection LSTM)**:** it uses a BiLSTM for which semantic relatedness of target with its context words are incorporated. A target vector is calculated by averaging the vectors of the words that compose the target. This method has optimal performance by using a simple average of the word embeddings of a target.
- **BiLSTM-CRF** (Chen et al., 2017)**:** it is a combination of BiLSTM and CRF layers to extract the targets related to the aspects. (Kooli and Pigneul, 2018) also uses BiLSTM and CRF for aspect extraction but only on French datasets.

Attention Mecanism. Opposite polarities may be detected when multiple aspects are considered. Thus, the use of only target information is not sufficient. The application of an attention mechanism can extract the association of important words denoting an aspect (Wang et al., 2016). It allows to capture the key part of a sentence in relation with a given aspect and to identify the associated sentiment.

Let us consider the following example: "The *menu* is *great* but *service* is a disaster". For the target *menu*, the word *great* will have a high weight and for the target *service*, the word *disaster* will have a high weight. Through attention mechanism, the network can learn the association of a positive sentiment for the *menu* and and a negative one for the *service*.

- **Attention-based LSTM with Aspect Embedding (ATAE-LSTM):** it strengthens the effect of target embeddings by appending it with each word embeddings and uses BiLSTM with attention to get the final representation for the detection of the aspects and the related sentiments. (Wang et al., 2016).
- **Interactive Attention Network (IAN):** (Ma et al., 2017) uses this model to learn the representations of the target and its context with BiLSTMs and attentions interactively, that generates the representations for targets and contexts with respect to each other.
- **Attentional Encoder Network (AEN):** (Song et al., 2019) do not use recurrence and employ attention based encoders for modeling context and aspect.

3 *CLC*: A NEW ARCHITECTURE FOR ABSA

In this section, we present our new model: *CLC*. It combines **C**NN, Bi**L**STM and **C**RF layers. Here, we describe each layer of the proposed neural network. We also motivate our approach and explain why each layer of the model is used. The global architecture of the proposed model is illustrated in Figure 4.

3.1 Embedding Layer

Embedding is used to represent each word in a vector with defined dimensions. The resulting vectors represent the projections in a continuous vector space. The position of a word is calculated as a function of the words in its context. The embedding for a word corresponds to the position of this word in the learned vector space. Nowadays, there are a lot of pre-trained word embeddings on corpus of millions of words like Word2Vec, GloVe or Wikipedia2Vec

3.2 CNN

We use CNN layer in order to get a word-level representation for each word of a phrase. (Kim, 2014) proved that a simple CNN gives very good performance for text classification with one layer of convolution on top of word vectors from an unsupervised neural language model. (Chiu and Nichols, 2016) also demonstrated that CNN is efficient for taking into account some morphological information like suffix or prefix of a word and encodes it into a neural representation. Figure 1 shows the different steps we used for the CNN layer in the *CLC* model.

3.3 LSTM

LSTM has been proposed to solve the vanishing/exploding gradient problem related to Recurrent Neural Network (RNN). A LSTM unit is composed of three gates: input, output and forget which can check the proportion of information to forget or to keep for the next step. Figure 2 shows in details a LSTM unit with its different components.

The formulas for updating a LSTM unit at time t are:

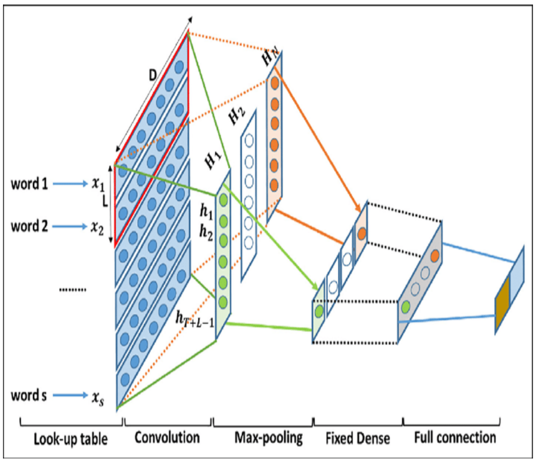

Figure 1: CNN on text (Nguyen et al. 2019).

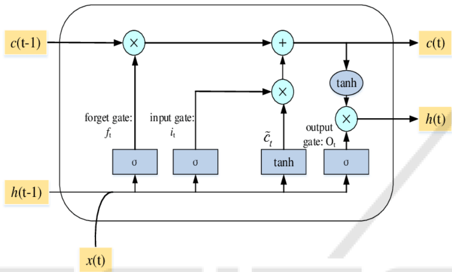

Figure 2: LSTM unit (Yuan et al. 2019).

$$i_t = \sigma(W_i h_{t-1} + U_i x_t + b_i)$$

$$f_t = \sigma(W_f h_{t-1} + U_f x_t + b_f)$$

$$\tilde{c}_t = \tan(W_c ht - 1 + U_c x_t + b_c)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t$$

$$o_t = \sigma(W_0 h_{t-1} + U_0 x_t + b_0)$$

$$h_t = o_t \odot \tan(c_t)$$

where σ is the element-wise sigmoïd function and $\odot$ is the element-wise product. x_t corresponds to the input vector (like word embedding) at time t, and h_t is the hidden state (or output) vector storing all the useful information at (and before) time t. U_i, U_f, U_c, U_o denote the weight matrices of different gates for input x_t, and W_i, W_f, W_c, W_o are the weight matrices for hidden state h_t. b_i, b_f, b_c b_o denote the bias vectors.

BiLSTM consists of two LSTM: one with the input in a forward direction, and the other in a backwards direction. BiLSTM allows to fully take into account the context of each word. Thus, we can learn the information from context which facilitates the aspect detection and the analysis of related sentiments. Figure 3 shows the BiLSTM architecture.

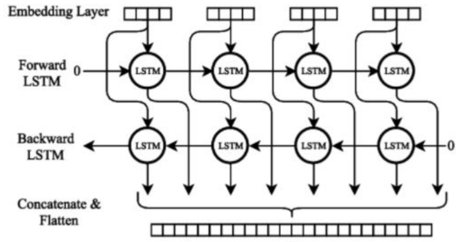

Figure 3: BiLSTM architecture (from S.Cornegruta and al, 2016.

3.4 CRF

CRF is used in statistical modelling for sequence labeling, which can be applied for aspect detection and ABSA.

The CRF layer with its constraints ensure that the predicted labels are valid. During the training, these constraints are learned automatically. Formally, let $z = \{z_1, \dots, z_n\}$ be a phrase with z_i, the vector for the i^{th} word and $y = \{y_1, \dots, y_n\}$ be the generic sequence of aspects for z. $\gamma(z)$ constitutes the set of possible aspects for z. CRF corresponds to a family of conditional probability $P(y|z;W;b)$ over all possible aspect sequences y given z and can be formulated as follows:

$$P(y|z;W;b) = \frac{\prod_{i=1}^{n} \psi_i(y_{i-1}, y_i, z)}{\sum_{y' \in \gamma(z)} \prod_{i=1}^{n} \psi_i(y'_{i-1}, y'_i, z)}$$

where $\psi_i(y', y, z) = exp(W_{y',y}^T z_i + b_{y',y})$ are potential functions, and $W_{y',y}^T$ and $b_{y',y}$ are the weight vector and bias corresponding respectively to label pair (y', y).

3.4.1 IOB Tagging

IOB (Inside, Outside, Begin) is a widely used tagging scheme. The B- tag corresponds to the beginning of an annotated chunk (target). The I- tag indicates that the tag is inside a chunk. Only a B- tag (or another I- tag for n-grams chunks) could precede an I- tag. The O- tag corresponds to the tokens that do not belong to any annotated chunks.

We can define the constraints of the CRF layer by using the IOB scheme. Hence, we can define some constraints such as:

- the first word of a sentence starts with a B- or an O- tag, and never with an I- tag;
- an O- tag followed by an I- tag is an invalid sequence;
- the first tag for a target is a B- tag and never an I- tag.

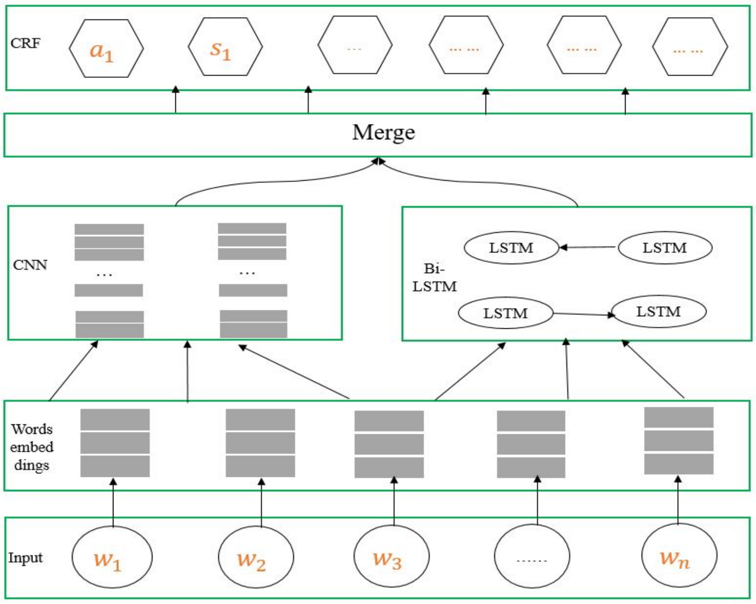

Figure 4: Architecture of *CLC*. We observe that only the word w_1 denotes an aspect. At the CRF layer, we predict a_1 as the aspect and s_1 as the associated sentiment.

Table 1: IOB Tagging.

meat	price	is	affordable
B_Meal#Price	I_Meal#Price	O	O

For example, if the sentence *"meat price is affordable"* is encoded with the IOB tagging scheme, we can have this labeling:
We notice in Table 1 that the target *meat price* is annotated with a B- and an I- tag for the aspect *Meal#Price*. It means that we consider here an unique target composed of two words and not two distinct targets (for the aspect *Meal#Price*).

3.5 Joint Aspect and Sentiment Detection

Many ABSA models used separate methods for the two following tasks:

- detection : detecting all the aspects;
- classification : classifying the polarity for each detected aspect.

In this paper, we propose a joint model able to detect the aspects and the associated sentiments at the same time.

Our *CLC* model is described in Figure 4. As mentioned earlier, the CNN layer is an effective approach to extract morphological information like the prefix or suffix of a word and encode it into neural representation. This allows a better word-level representation. The input used for the CNN layer is the word embeddings. Then, the CNN layer acts as a character embedding. We feed also a BiLSTM layer with the word embeddings. The output vectors from BiLSTM and CNN are merged and finally, the merged vectors are given as input to the CRF layer in order to predict the aspects and the related sentiments.

Let W and T be the two inputs that we used for fitting the model: W corresponds to the words and T to the corresponding tags. In Table 2 we can see an example of W and T given a customer review.

Table 2: Inputs for an example of a customer review.

W	Price	was	affordable
T	B_Meal#Price	O	O

In order to jointly detect aspects and sentiments, we insert an index for polarity just after the corresponding aspect in W and T and we give these new inputs to the model for the training (see Table 3). This allows *CLC* to detect aspects and sentiments at the same time.

Table 3: Inputs for joint aspect and sentiment detection.

W	Price	[sentiment]	was	affordable
T	B_Meal#Price	positive	O	O

For these new inputs dedicated to joint detection, we also need to encode the sentiments. Table 4 shows the encoding we use for positive, negative or neutral sentiment. If there is n aspects in the dataset, we can

```xml
<Review rid="g187079-d3590623-r272464202">
    <sentences>
        <sentence id="g187079-d3590623-r272464202:0">
            <text>Viande de boeuf façon chewing gum, frites couleur marron foncé molles et baignant encore dans leur huile.</text>
            <Opinions>
                <Opinion target="Viande de boeuf" category="FOOD#QUALITY" polarity="negative" from="0" to="15"/>
                <Opinion target="frites" category="FOOD#QUALITY" polarity="negative" from="35" to="41"/>
            </Opinions>
        </sentence>

        <sentence id="g187079-d3590623-r272464202:3">
            <text>Nous avons rarement aussi mal mangé à Bordeaux.</text>
            <Opinions>
                <Opinion target="NULL" category="FOOD#QUALITY" polarity="negative" from="0" to="0"/>
            </Opinions>
        </sentence>
    </sentences>
</Review>
```


Figure 5: Example of review from French SemEval2016 dataset on restaurants. We note that the aspect is tagged by *category*. For implicit aspect the target is *NULL*, because no words in the sentence is directly referring to the aspect.

observe in table 4 that the positive sentiment is encoded by *n+1*, the negative sentiment is encoded by *n+2* and the neutral one by *n+3*. It is an arbitrary choice in order to avoid confusion with the encoding of the vocabulary words.

Table 4: Encoding of the sentiments for joint detection with *n* being the number of aspects.

Sentiment	Encoding
positive	n + 1
negative	n + 2
neutral	n + 3

4 EXPERIMENT

In this section, we show how our *CLC* model is well adapted for aspect and sentiment detection on French language. We give details about the dataset we used and we demonstrate that *CLC* outperforms the existing methods for French ABSA.

4.1 SemEval2016 Dataset

French Dataset. In our experiment, we use the French SemEval2016 annotated dataset about restaurants (Apidianaki et al., 2016). It contains 337 reviews for the training set and 120 reviews for the test set. The dataset is annotated with targets, aspects and sentiments (Figure 5). Since each review can be composed of several sentences, the dataset actually contains 1669 annotated sentences for training set and 696 for the test set. It incorporates 12 categories of aspects and 3 types of sentiments (positive, negative and neutral). In Figure 5, we give an extract of the dataset.

English Dataset. In order to confirm that *CLC* is adapted for other languages, we also test our model with a SemEval2016 annotated dataset in English. In the training set of this dataset, there are 2000 sentences coming from 350 reviews. The test set contains 676 sentences from 90 reviews.

4.2 Training Details

Before training our model we perform some preprocessing steps such as removing the stop words and doing lemmatization in order to improve the *CLC* performance. We use publicly available Wikipedia2Vec with 300-dimensional embeddings trained on words from Wikipedia.

For CNN Layer, we used 50 filters with window length equal to 4. The BiLSTM layer network size is 300.

4.3 Results

Table 5 shows a comparison of our *CLC* method with the state-of-the-art models. Since we are interested in French language, we only consider methods that give the highest performances on the French SemEval datasets. We present the F-score for aspect detection using the French SemEval2016 dataset on restaurants. Generally, the reference methods only focus on aspect detection or give separate results for the aspects and the sentiments.

We can observe in Table 5, that our *CLC* model outperforms the state-of-the-art models when we focus on ACD. Therefore, *CLC* shows great interest when it comes to the French language. Thus, *CLC* can be used to fill the gap that exists about efficient methods for aspect detection in French language.

During our study, we observed that using a CRF layer instead of a softmax activation improves the F-score from 74.2% to 77.2%. The reason is that the CRF layer allows to predict the label using neighbouring tagging predictions that eliminates some false predictions.

As mentioned in section 3.5, our *CLC* model can

Table 5: Aspect detection Comparison for French SemEval-2016 dataset about restaurants.

	F1-Score
(Kumar et al., 2016)	57.8
(Ruder et al., 2016)	61.2
(Brun et al., 2016)	61.2
(Poria et al., 2016)	67.06
(Chen et al., 2017)	73.50
(Kooli and Pigneul, 2018)	70.97
CLC	**77.2**

also be deployed for joint aspect and sentiment detection. We observe that when we use *CLC* as a joint model, the F1-score for the couple (aspect, sentiment) reaches 88.7%. The reason is that *CLC* learns well the correlation between the aspect and the sentiment. Thus, more aspects are well detected.

***CLC* Applied on Other Datasets.** In order to prove the generalization capability of *CLC*, we apply it on an unseen dataset which is the french SemEval2016 dataset about Museum. The *Museum* dataset contains 162 french annotated reviews with 668 sentences for testing. *CLC* achieves great performances in this context with a a F1-score equal to 69.97%

To prove that *CLC* is also adapted for English language, we tested it on a SemEval2016 English dataset on restaurants. We achieve in this case great performance with a F1-score equal to 80.14% and shows that our model can be used for ABSA in multiple languages.

5 CONCLUSION

In this paper we propose a new model named *CLC* for ABSA and more precisely for aspect and sentiment detection.

Several efficient methods have been proposed for English language in literature. However, only a few work has been done in French due to the lack of resources. Thus, the main contribution of our work is that we propose a new model that gives high performance on the French language. *CLC* is also well adapted for other language as we proved it for English.

The strength of the proposed model is its ability to jointly detect the aspects and the associated sentiment at the same time and thus fully exploit the correlation between them.

For the future works, it could be interesting to add an attention mechanism in *CLC* which may improve its performance. There is also a great need to develop unsupervised models for ABSA to get rid off the dependency on annotations. This would open new perspectives to explore since there is a huge amount of unannotated data available on the web (e.g. customer reviews on TripAdvisor).

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)*.

Brun, C. and Nikoulina, V. (2018). Aspect based sentiment analysis into the wild. In *Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, WASSA@EMNLP 2018, Brussels, Belgium, October 31, 2018*.

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@NAACL-HLT 2016, San Diego, CA, USA, June 16-17, 2016*.

Chen, T., Xu, R., He, Y., and Wang, X. (2017). Improving sentiment analysis via sentence type classification using bilstm-crf and CNN. *Expert Syst. Appl.*, 72:221–230.

Chiu, J. P. C. and Nichols, E. (2016). Named entity recognition with bidirectional lstm-cnns. *Trans. Assoc. Comput. Linguistics*, 4:357–370.

Dong, L., Wei, F., Tan, C., Tang, D., Zhou, M., and Xu, K. (2014). Adaptive recursive neural network for target-dependent twitter sentiment classification. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014, June 22-27, 2014, Baltimore, MD, USA, Volume 2: Short Papers*.

Hamdan, H., Bellot, P., and Béchet, F. (2015). Lsislif: CRF and logistic regression for opinion target extraction and sentiment polarity analysis. In *Proceedings of the 9th International Workshop on Semantic Evaluation, SemEval@NAACL-HLT 2015, Denver, Colorado, USA, June 4-5, 2015*.

Kim, Y. (2014). Convolutional neural networks for sentence classification. *CoRR*, abs/1408.5882.

Kiritchenko, S., Zhu, X., Cherry, C., and Mohammad, S. (2014). Nrc-canada-2014: Detecting aspects and sentiment in customer reviews. In *Proceedings of the 8th International Workshop on Semantic Evaluation, SemEval@COLING 2014, Dublin, Ireland, August 23-24, 2014*.

Kooli, N. and Pigneul, E. (2018). Analyse de sentiments à base d'aspects par combinaison de réseaux profonds : application à des avis en français (A combination of deep learning methods for aspect-based sentiment analysis : application to french reviews). In *Actes de la Conférence TALN. CORIA-TALN-RJC 2018 - Volume 1, Rennes, France, May 14-18, 2018*.

Kumar, A., Kohail, S., Kumar, A., Ekbal, A., and Biemann, C. (2016). IIT-TUDA at semeval-2016 task 5: Beyond sentiment lexicon: Combining domain dependency and distributional semantics features for aspect based sentiment analysis. In *Proceedings of the 10th International Workshop on Semantic Evaluation, SemEval@NAACL-HLT 2016, San Diego, CA, USA, June 16-17, 2016*.

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 2017, Melbourne, Australia, August 19-25, 2017*.

Machacek, 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@NAACL-HLT 2016, San Diego, CA, USA, June 16-17, 2016*.

Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., Al-Smadi, M., Al-Ayyoub, M., Zhao, Y., Qin, B., Clercq, O. D., Hoste, V., Apidianaki, M., Tannier, X., Loukachevitch, N. V., Kotelnikov, E. V., Bel, N., Zafra, S. M. J., and Eryigit, G. (2016). Semeval-2016 task 5: Aspect based sentiment analysis. In *Proceedings of the 10th International Workshop on Semantic Evaluation, SemEval@NAACL-HLT 2016, San Diego, CA, USA, June 16-17, 2016*.

Poria, S., Cambria, E., and Gelbukh, A. F. (2016). Aspect extraction for opinion mining with a deep convolutional neural network. *Knowl. Based Syst.*, 108:42–49.

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@NAACL-HLT 2016, San Diego, CA, USA, June 16-17, 2016*.

Song, Y., Wang, J., Jiang, T., Liu, Z., and Rao, Y. (2019). Attentional encoder network for targeted sentiment classification. *CoRR*, abs/1902.09314.

Tang, D., Qin, B., Feng, X., and Liu, T. (2016a). Effective lstms for target-dependent sentiment classification. In *COLING 2016, 26th International Conference on Computational Linguistics, Proceedings of the Conference: Technical Papers, December 11-16, 2016, Osaka, Japan*.

Tang, D., Qin, B., and Liu, T. (2016b). Aspect level sentiment classification with deep memory network. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016, Austin, Texas, USA, November 1-4, 2016*.

Thet, T. T., Na, J., and Khoo, C. S. G. (2010). Aspect-based sentiment analysis of movie reviews on discussion boards. *J. Inf. Sci.*, 36(6):823–848.

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, EMNLP 2016, Austin, Texas, USA, November 1-4, 2016*.

Wu, H., Gu, Y., Sun, S., and Gu, X. (2016). Aspect-based opinion summarization with convolutional neural networks. In *2016 International Joint Conference on Neural Networks, IJCNN 2016, Vancouver, BC, Canada, July 24-29, 2016*, pages 3157–3163. IEEE.

Young, T., Hazarika, D., Poria, S., and Cambria, E. (2018). Recent trends in deep learning based natural language processing [review article]. *IEEE Comput. Intell. Mag.*, 13(3):55–75.