Document Knowledge Transfer for Aspect-Based Sentiment Classification Using a Left-Center-Right Separated Neural Network with Rotatory Attention

Emily Fields, Gonem Lau, Robbert Rog, Alexander Sternfeld, and Flavius Frasincar[0000−0002−8031−758X] (✉)

Erasmus University of Rotterdam, Burgemeester Oudlaan 50, 3062 PA Rotterdam, the Netherlands
{505456ef,500202gl,492751rr,492825as}@student.eur.nl, frasincar@ese.eur.nl

Abstract. Hybrid Aspect-Based Sentiment Classification (ABSC) methods make use of domain-specific, costly ontologies to make up for the lack of available aspect-level data. This paper proposes two forms of transfer learning to exploit the plenteous amount of available document data for sentiment classification. Specifically, two forms of document knowledge transfer, pretraining (PRET) and multi-task learning (MULT), are considered in various combinations to extend the state-of-the-art LCR-Rot-hop++ model. For both the SemEval 2015 and 2016 datasets, we find an improvement over the LCR-Rot-hop++ neural model. Overall, the pure MULT model performs well across both datasets. Additionally, there is an optimal amount of document knowledge that can be injected, after which the performance deteriorates due to the extra focus on the auxiliary task. We observe that with transfer learning and L1 and L2 loss regularisation, the LCR-Rot-hop++ model is able to outperform the HAABSA++ hybrid model on the (larger) SemEval 2016 dataset. Thus, we conclude that transfer learning is a feasible and computationally cheap substitute for the ontology step of hybrid ABSC models.

Keywords: LCR-Rot-hop++ · Transfer Learning · Pretraining · Multi-Task Learning

1 Introduction

In the pre-Web era, it was often difficult for companies to gauge the opinions of their large customer bases. While the increasing popularity of the Web provided a virtually inexhaustible source of data, machine learning methods had to be developed for extracting insights from this information. One particularly interesting insight is to extract a sentiment from a segment of text, for example a review. This is what drove the development of Sentiment Analysis in the field of Natural Language Processing (NLP) [9].

This paper specifically considers Aspect-Based Sentiment Analysis (ABSA). ABSA consists of two steps: Aspect Detection (AD) and Aspect-Based Sentiment Classification (ABSC). AD is the task of finding an aspect, such as price, quality, or service of an entity, within a text or review. This paper will focus on ABSC exclusively, which involves determining the sentiment of a given aspect within a given sentence [1, 14]. It is common practice to divide sentiment into three classes: positive, neutral, and negative.

Traditionally, dictionary-based approaches such as that in [7] have been used, but with the rise of deep learning and the ever-increasing computational power of modern machines, a range of new techniques for ABSC have become available. Of the basic "deep" models, the Bidirectional Long Short-Term Memory Network (BiLSTM) at first appeared to be the most promising, as illustrated in [5]. Over the years, however, more sophisticated BiLSTM methods have been developed. One of these is the Left-Center-Right Separated BiLSTM with Rotatory Attention (LCR-Rot) [20], which utilises separately trained BiLSTM networks for the context to the left of the aspect, the aspect itself, and the context to the right of the aspect, and has been found to outperform previously proposed LSTM-variations [20]. Even more recently, the LCR-Rot model has been extended with respect to both the attention mechanism and the word embeddings. Namely, the LCR-Rot-hop model presented in [19] iteratively applies the attention mechanism, while the LCR-Rot-hop++ model proposed in [18] builds on this to include hierarchical attention and deep contextual word embeddings.

Over the past years, a collection of techniques called Transfer Learning (TL) has surged in popularity as a method to improve the performance of machine learning methods [10]. More formally, TL involves training a model on auxiliary tasks to improve the performance of the main task. This is particularly interesting when there is few data available for our task at hand. One such method is MULTi-task learning (MULT), where a model is trained on two tasks simultaneously, as applied in the widely used language model called Bidirectional Encoder Representation from Transformers (BERT) [4]. An alternative method is PRE-Training (PRET), which involves first learning an auxiliary task after which the model is trained for the main task. The latter step is called Fine-Tuning (FT), meaning a TL model is trained once more on just the main task.

A lack of training data in the same domain as the test data is a persistent issue in machine learning [10]. In ABSC, this is reflected by the limited availability of aspect-level data. As there is more annotated sentiment data available at a document level, i.e., review texts with star ratings, this information can be exploited using TL techniques. [6], for example, showed an improvement in the performance of ABSC in BiLSTMs when PRET and MULT are utilised. We consider four approaches for document knowledge transfer in the state-of-the-art LCR-Rot-hop++ neural model, inspired by [6]. The first approach is PRET+FT, in which the model is first pretrained on the document knowledge and then fine-tuned. The second is MULT, where the sentiment of a document and of an aspect are determined simultaneously. While the method proposed in [6] does not include a regularisation term, we extend the approach by including a regularisation term in the loss function as in [19]. The third method is a combination of both PRET and MULT, called PRET+MULT, in which the model is first pretrained at a document level on part of the data, before MULT is applied to the rest of the data. Last, in the fourth and fifth approaches, we develop new methods that incorporate FT into the TL approach, in two models called MULT+FT and PRET+MULT+FT.

The present work extends the existing literature by implementing document knowledge transfer on the state-of-the-art LCR-Rot-hop++ neural model, with the aim to further improve its accuracy. Moreover, different L1 and L2 regularisation terms are combined to improve upon the previous works. The Python source code of our models can be found at https://github.com/Gogonemnem/LCR-Rot-hop-plus-plus-TL.

The paper continues as follows. First, the related works and their results are discussed in more detail in Sect. 2, after which the data is illustrated in Sect. 3. Subsequently, the methodology is presented in Sect. 4. Thereafter, the results are compared with those obtained in the previous literature in Sect. 5. Last, conclusions with the main findings and suggestions for future research are presented in Sect. 6.

2 Related Work

The performance of the LCR-Rot-hop++ model depends on the scale of the available training data, as limited training data can lead to a lower accuracy. Ideally, one would use aspect-level data, as the model is used for sentiment analysis at the aspect level. However, the availability of annotated aspect-level data is limited [6,10]. To illustrate, both [18] and [19] use the standard SemEval 2015 [12] and SemEval 2016 [11] datasets, which are relatively small. Due to the limited availability of aspect-level data, the LCR-Rot-hop++ model may not reach its full potential. To overcome this issue, one can consider coarser data, such as document-level or sentence-level data. There is an abundance of this type of data, for instance Yelp reviews [16].

Document knowledge transfer can be motivated from three perspectives: human learning, pedagogy, and machine learning [13]. From the point of view of human learning, it is clear that we frequently use knowledge acquired from learning related tasks when learning a new task. Equally, from a pedagogical perspective, we often learn the foundations first, before using this knowledge to learn more complex skills. Last, document knowledge transfer improves generalisation by introducing an inductive bias, which creates a preference for hypotheses that explain more than one task [2]. In this paper, we investigate which method for document knowledge transfer performs best, specifically, we consider combinations of PRETraining (PRET), MULTi-task learning (MULT), and Fine-Tuning (FT).

PRET. Pretraining is the act of training a model on a task semantically related to your target task, prior to training for your target task [6,13]. This technique has shown great success in language models such as BERT in [4]. BERT was trained to perform two tasks which helped the model understand language, after which it can be fine-tuned for a wider variety of language tasks. As shown in [6], pretraining a BiLSTM on document-level data improved the results obtained on an aspect level.

MULT. In contrast to PRET, when using MULT the model is trained for the target task and the semantically related task simultaneously [13]. The purpose of this is to improve generalisation, which might lead to more effective knowledge transfer. For example, [15] demonstrated that multi-task learning is able to produce good word embeddings.

FT. In PRET and MULT, one uses the semantically-related task to improve performance on the target task. With FT, one only trains on the target task. Therefore, it is necessary in combination with PRET, but optional with MULT. To illustrate, the BERT language model is first trained using a MULT approach on general language tasks, after which it can be trained for specific tasks using an FT approach [4].

3 Data

The datasets used for ABSC are the SemEval 2015 [12] and SemEval 2016 [11] datasets. Specifically, our analysis focuses on restaurant reviews. Each review consists of one or more sentences, and each sentence contains the sentiment on one or more aspects. The sentiment can either be positive, neutral, or negative. In our research we focus on explicit aspects, which means that the aspect is present in the sentence. Figure 1 shows an example sentence from the SemEval 2016 dataset in the XML markup language. This example shows that, in a review, multiple aspects can be present and the sentiment towards different aspects may differ. Table 1 gives the descriptive statistics of the SemEval 2015 and the SemEval 2016 data sets. One can notice that there are relatively few neutral reviews. Furthermore, in most data the positive class is the majority class, except for the test data of the SemEval 2015 [12] dataset. The 2015 dataset is a subset of the 2016 dataset, and is noticeably smaller.

We use a document-level dataset from Yelp2014 [17] for pretraining, as it matches the domain of our aspect-level data: restaurants. However, these reviews are classified on a 5-point scale. Therefore, the reviews will be labeled in the following way: reviews with ratings > 3, $= 3$, and < 3 are labeled as positive, neutral, and negative, respectively. Similar to [6], a balanced sample of 30,000 is extracted from the dataset to obtain our pretraining corpus. As Table 1 shows, there is a significant lack of neutral examples in the aspect-level data. Therefore, the balancing of the pretraining corpus allows the model to see an ample amount of documents for each category. To make our data suitable for multi-task learning, each aspect-level data point is paired with a random document from our

```xml
<sentence id="Z#3:0">
    <text>Excellent food, although the interior could use some help.</
        text>
    <Opinions>
        <Opinion target="food" category="FOOD#QUALITY" polarity="positive"
             from="10" to="14"/>
        <Opinion target="interior" category="AMBIENCE#GENERAL" polarity="
            negative" from="29" to="37"/>
    </Opinions>
</sentence>
```

Fig. 1. A sentence from the SemEval 2016 dataset.

sample. As there are many more documents available than aspects, we upsample our aspect-level data with a factor of three. This value was chosen based on intuition, as too little upsampling will not allow us to exploit many documents, whereas too much upsampling will likely lead the model to overfit due to the duplicates in the aspect data.

Table 1. Descriptive statistics of the SemEval 2015 and SemEval 2016 datasets, split into training and test data.

Dataset	Positive		Neutral		Negative		Total
	Freq	%	Freq	%	Freq	%	Freq
SemEval-2015 training data	963	75.3	36	2.8	280	21.9	1279
SemEval-2015 test data	354	34.7	38	6.3	208	59.0	600
SemEval-2016 training data	1321	70.1	73	3.9	490	26.0	1884
SemEval-2016 test data	487	74.4	32	4.9	136	20.8	655

4 Methodology

This section discusses the methodology used to obtain the results. First, the two methods for document knowledge transfer (PRET and MULT) are elaborated upon, along with the various combinations in which they are applied. Then, the experimental setup for obtaining the results is given.

4.1 Knowledge Transfer

This paper considers several different approaches for document knowledge transfer. Each approach consists of one or more of the following building blocks: PRET, MULT, and FT. In this section, each building block is described separately. Furthermore, Fig. 2 displays which compositions of PRET, MULT, and FT are used in this work. In the notation of each building block, we consider two tasks τ_1 and τ_2. Let τ_2 be our task of interest, namely ABSC. In contrast, τ_1 is sentiment classification at a document level, which is semantically related to our main task. Therefore, teaching our model to execute τ_1 will enlighten it with knowledge that can be used for better executing τ_2.

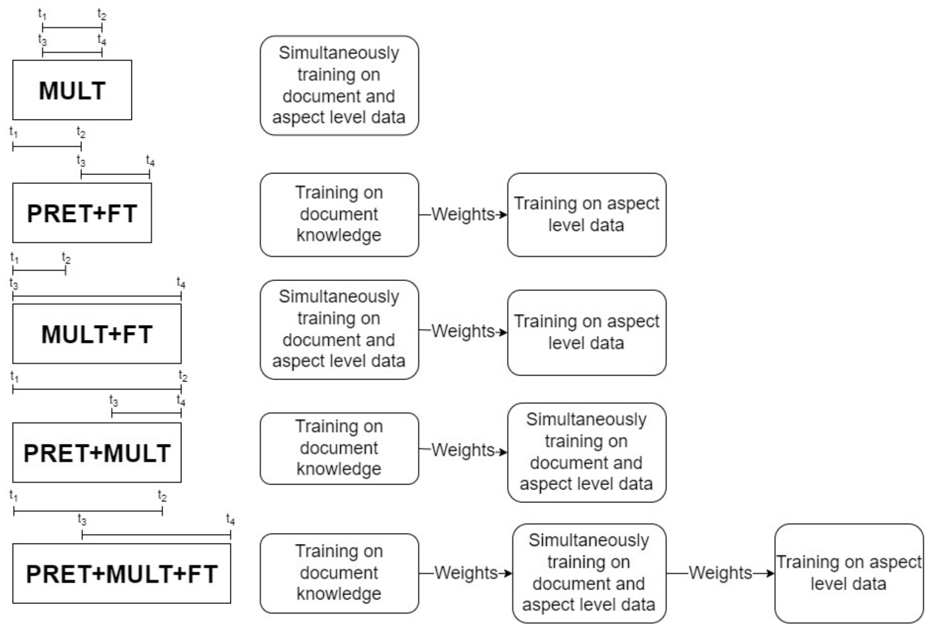

Fig. 2. An overview of the different document knowledge transfer approaches. The target task is executed in the time interval $[t_3, t_4]$, whereas the semantically related task is executed in the time interval $[t_1, t_2]$

PRET. In the pretraining stage only τ_1 is executed, which trains the model for sentiment analysis at a document level. Specifically, the documents are put through the left, center, and right BiLSTM, after which the final hidden layers of all words are pooled (averaged). The pooled hidden layers of the three BiLSTMs are concatenated and fed into a classification layer. The aim of this stage is to pretrain the BiLSTMs, as it is expected that the BiLSTM weights obtained in the PRET stage transfer the information from the document-level sentiment classification to improve the accuracy at the aspect level.

MULT. During the multi-task stage, tasks τ_1 and τ_2 are executed simultaneously. In this approach, the three BiLSTMs are trained simultaneously on the document-level data and on their corresponding part of the aspect-level data (left, target, or right). Each aspect-level data point is paired to a document-level data point. Thus, the embedding layer and the three BiLSTMs in the LCR-Rot-hop++ model are shared for τ_1 and τ_2. The document-level data is processed the same as in the PRET stage. For the aspect-level data, the outputs from the BiLSTMs are directed to the corresponding attention mechanism, which finally leads to probabilities regarding the aspect-based sentiment.

The parameters are set by minimising the loss function below.

$$L = J + \lambda U + \omega \|\Theta\|_1 + \Omega \|\Theta\|^2 \tag{1}$$

In this loss function, J is the mean loss per training batch corresponding to our primary task τ_2. Likewise, U is the mean loss per training batch corresponding to our secondary task τ_1. The loss U is weighted with a parameter $\lambda \in (0, 1)$, which can be interpreted as the importance of τ_1 for performing τ_2. Last, ω and Ω denote the weights of the L1 and L2 regularisation terms, respectively. The L1 regularisation considers the absolute value of the coefficients, whereas the L2 regularisation considers the squared value of the coefficients.

FT. The fine-tuning stage can be used as the final stage for training a model. In the FT stage, only τ_2 is executed. In this context, this means that only ABSC is performed. The goal of this stage is to tweak the model, such that it performs best for the target task. Whereas previous stages have taught the model more general knowledge, the FT stage aims at preparing the model solely for the target task.

4.2 Experimental Setup

To verify the added value of the TL approaches, we test all combinations as presented in the previous section and compare their performance to the benchmark model without document knowledge transfer. The following section describes in more detail how we find the best models for each combination.

Hyperparameter Tuning. Hyperband is used to find the optimal hyperparameters [8]. As hypertuning all models over all stages is computationally infeasible, a heuristic is used for setting the hyperparameters. Namely, for each dataset, the optimal hyperparameters for the MULT model and the FT model are computed. These hyperparameters are generalised over all building blocks of the model. Models which use hyperparameters from the tuned FT model are referred to as FT-based models. Models which use hyperparameters from the tuned MULT model are referred to as MULT-based models. Thus, each approach in Fig. 2 is executed using both the FT hyperparameters and the MULT hyperparameters. For the FT-based PRET+MULT+FT model, λ has not been optimised in the hypertuning. Hence, the λ from MULT tuning is generalised to this model as well. Note that we do not run an FT-based model for MULT nor PRET+MULT, nor a MULT-based model for PRET+FT, as these parameter and TL approach combinations are likely to be suboptimal. We use 80% of the training data to optimize the loss function and the remaining 20% as validation to select the best hyperparameters.

Model Training. Early stopping is applied to determine the number of training epochs with different levels of patience for the stages. This means that when the performance on the validation set has not increased during the patience epochs, the epochs after the current best epoch, training is stopped and the optimal model weights are restored. For the PRET stage, the performance measure is the validation loss (categorical cross-entropy). The PRET corpus is large compared to the aspect level data, so for computational efficiency a relatively low patience of 3 is chosen here. Similarly, for the MULT stage, we use early stopping with respect to the combined validation loss described in [6]. We allow a higher patience of 10 as the per epoch time is considerably lower, making it more affordable. Last, for the FT stage, early stopping is done with respect to the validation accuracy, as this is our measure of interest. Again, a patience of 10 is used. The loss functions in all stages, including the benchmark LCR-Rot-hop++, are regularised to prevent overfitting. Both L1 and L2 regularisation are used, the weights of which are optimised by the aforementioned hyperband for both FT-based and MULT-based models.

Model Evaluation. We evaluate the various approaches using the out-of-sample accuracy measure. This measure allows us to see, after training, how often a model correctly predicts the sentiment of an aspect. We note that this measure weights the performance for each sentiment class according to how many observations there are for each sentiment, meaning it does not heavily penalise poor performance in a small sentiment class. As we observe in our data that there are very few neutral observations compared to positive and negative observations, we acknowledge that poor model performance when predicting a neutral sentiment might not be strongly reflected in the accuracy.

5 Results

Table 2 shows the results of the benchmark LCR-Rot-hop++ model with different hyperparameters and different combinations of document knowledge transfer approaches, for the data of SemEval 2015 and SemEval 2016. All losses, including that of the benchmark LCR-Rot-hop++ model without document knowledge transfer, are regularised using the L1 and L2 regularisation terms, allowing for a fair comparison. We find that several TL models outperform the benchmark LCR-Rot-hop++ model, for both datasets, suggesting there is added value in incorporating document knowledge transfer in the base model.

For the SemEval 2015 dataset, all models with TL outperform the benchmark model. The largest improvement in accuracy, 6.50 percentage points, is observed for the MULT model. For the SemEval 2016 dataset, on the other

Table 2. Results of LCR-Rot-hop++ with various forms of document knowledge transfer, alongside the benchmark model without document knowledge transfer, for the SemEval 2015 and SemEval 2016 datasets.

Settings	Accuracy	
	SemEval 2015	SemEval 2016
Benchmark model		
LCR-Rot-hop++	74.00%	86.87%
FT-based models		
MULT+FT	77.00%	85.95%
PRET+FT	78.00%	88.70%
PRET+MULT+FT	79.67%	86.56%
MULT-based models		
MULT	80.50%	88.63%
MULT+FT	74.50%	87.18%
PRET+MULT	76.67%	85.04%
PRET+MULT+FT	77.67%	86.87%

Note. The FT-based models are constructed using the optimal hyperparameters from a model with only the FT stage, as is the LCR-Rot-hop++ benchmark model. The MULT-based models use the optimal hyperparameters for a pure MULT model.

hand, several models with TL do not outperform the benchmark, namely FT-based MULT+FT and PRET+MULT+FT, and MULT-based PRET+MULT and PRET+MULT+FT. Still, the remaining models do outperform the benchmark, with the biggest improvements observed for the PRET+FT and MULT models, which exceed the accuracy of the benchmark model by 1.83 and 1.76 percentage points, respectively. The differences in performance for this dataset are smaller, likely because it is larger and more balanced. Given that TL aims to handle limited data availability and data imbalance by supplying the model with additional examples of a similar task, it is indeed to be expected that these approaches have greater impact in the 2015 dataset.

Based on the accuracy measures, we conclude that the analysed TL models can boost the accuracy of the existing LCR-Rot-hop++ model. Overall, the MULT model performs best, as it leads to the greatest improvements in accuracy across the two datasets. In comparison to the existing state-of-the-art HAABSA++ model, we observe that our MULT model outperforms HAABSA++ for the SemEval 2016 dataset, by 1.63 percentage points, but performs slightly worse for the 2015 dataset, by 1.2 percentage points.

One plausible reason for MULT outperforming PRET approaches is catastrophic forgetting [3]. Knowledge learned in the PRET stage might be forgotten when the model is retrained on the main task. In MULT, the main and auxiliary tasks are learned simultaneously, making the document knowledge more recent and prevalent. As shown in [3], multi-task learning provides a solution to catastrophic forgetting.

6 Conclusion

ABSC models are constrained due to the limited availability of aspect-level training data. In this paper, we aim to overcome this limitation by using document-level training data to train the state-of-the-art LCR-Rot-hop++ model. The results show that the most successful transfer learning approach is multi-task learning, particularly when faced with a small and imbalanced dataset such as SemEval 2015. Likely, multi-task learning outperforms the pretraining approach due to catastrophic forgetting; document knowledge acquired in pretraining is partly forgotten when the model is retrained on aspects. Multi-task learning solves this problem by fitting on the main and auxiliary task simultaneously, preventing this type of forgetting [3].

Our best approach, the MULT model, yields a 6.5 percentage points increase relative to the state-of-the-art LCR-Rot-hop++ model with L1 and L2 regularisation for the SemEval 2015 dataset, as well as a 1.76 percentage point increase for the SemEval 2016 dataset. Furthermore, this model outperforms the HAABSA++ model for the SemEval 2016 dataset. Therefore, we conclude that the inclusion of L1 and L2 regularisation terms along with the MULT method of document knowledge transfer can under certain circumstances effectively compensate for the exclusion of an ontology reasoning. Hence, this updated model can serve as a computationally cheaper alternative to existing hybrid models, without any significant loss in performance.

A suggestion for future research is to investigate different deep learning architectures for incorporating document knowledge transfer. One example of a different architecture is adding a shared BiLSTM layer below the LCR-Rot-hop++ model, instead of sharing the left, middle, and right BiLSTM. Furthermore, future research can investigate models that exploit sentence- or paragraph-level knowledge, besides document-level knowledge.

References

1. Brauwers, G., Frasincar, F.: A Survey on Aspect-Based Sentiment Classification. ACM Computing Surveys **55**(4), 65:1–65:37 (2023)
2. Caruana, R.: Multitask Learning: A Knowledge-Based Source of Inductive Bias. In: 10th International Conference on Machine Learning (ICML 1993). pp. 41–48. Morgan Kaufmann (1993)
3. Chen, S., Hou, Y., Cui, Y., Che, W., Liu, T., Yu, X.: Recall and Learn: Fine-tuning Deep Pretrained Language Models with Less Forgetting. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP 2020). ACL (2020)
4. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In: 17th Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2019). pp. 4171–4186. ACL (2019)
5. Graves, A., Schmidhuber, J.: Framewise Phoneme Classification with Bidirectional LSTM and other Neural Network Architecture. Neural Networks **18**(5-6), 602–610 (2005)

6. He, R., Lee, W.S., Ng, H.T., Dahlmeier, D.: Exploiting Document Knowledge for Aspect-level Sentiment Classification. In: 56th Annual Meeting of the Association for Computational Linguistics (ACL 2018). pp. 579–585. ACL (2018)
7. Hu, M., Liu, B.: Mining and Summarizing Customer Reviews. In: 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2004). pp. 168–177. ACM (2004)
8. Li, L., Jamieson, K., DeSalvo, G., Rostamizadeh, A., Talwalkar, A.: Hyperband: A Novel Bandit-Based Approach to Hyperparameter Optimization. Journal of Machine Learning Research **18**(1), 6765–6816 (2018)
9. Liu, B.: Sentiment Analysis: Mining Opinions, Sentiments, and Emotions. Cambridge University Press, 2nd edn. (2020)
10. Pan, S.J., Yang, Q.: A Survey on Transfer Learning. IEEE Transactions on Knowledge and Data Engineering **22**(10), 1345–1359 (2010)
11. 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., Eryiğit, G.: SemEval-2016 Task 5: Aspect Based Sentiment Analysis. In: 10th International Workshop on Semantic Evaluation (SemEval 2016). pp. 19–30. ACL (2016)
12. Pontiki, M., Galanis, D., Papageorgiou, H., Manandhar, S., Androutsopoulos, I.: SemEval-2015 Task 12: Aspect Based Sentiment Analysis. In: 9th International Workshop on Semantic Evaluation (SemEval 2015). pp. 486–495. ACL (2015)
13. Ruder, S.: Neural Transfer Learning for Natural Language Processing. PhD thesis, National University of Ireland, Galway (2019)
14. Schouten, K., Frasincar, F.: Survey on Aspect-Level Sentiment Analysis. IEEE Transactions on Knowledge and Data Engineering **28**(3), 813–830 (2016)
15. Subramanian, S., Trischler, A., Bengio, Y., Pal, C.J.: Learning General Purpose Distributed Sentence Representations via Large Scale Multi-task Learning. In: 6th International Conference on Learning Representation (ICLR 2018). OpenReview.net (2018)
16. Tang, D., Qin, B., Liu, T.: Learning Semantic Representations of Users and Products for Document Level Sentiment Classification. In: 53rd Meeting of the Association for Computational Linguistics (ACL 2015). pp. 1014–1023. ACL (2015)
17. Tang, D., Qin, B., Liu, T.: Aspect Level Sentiment Classification with Deep Memory Network. In: 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP 2016). pp. 214–222. ACL (2016)
18. Trușcă, M.M., Wassenberg, D., Frasincar, F., Dekker, R.: A Hybrid Approach for Aspect-Based Sentiment Analysis Using Deep Contextual Word Embeddings and Hierarchical Attention. In: 20th International Conference on Web Engineering (ICWE 2020). LNCS, vol. 12128, pp. 365–380. Springer (2020)
19. Wallaart, O., Frasincar, F.: A Hybrid Approach for Aspect-Based Sentiment Analysis Using a Lexicalized Domain Ontology and Attentional Neural Models. In: 16th Extended Semantic Web Conference (ESWC 2019). LNCS, vol. 11503, pp. 363–378. Springer (2019)
20. Zheng, S., Xia, R.: Left-Center-Right Separated Neural Network for Aspect-based Sentiment Analysis with Rotatory Attention. arXiv preprint arXiv:1802.00892 (2018)