A Hybrid Approach for Aspect-Based Sentiment Analysis Using a Double Rotatory Attention Model

Guangyao Zhou*
Jingyi Cheng
Flavius Frasincar

Econometric Institute,
Erasmus University Rotterdam,
P.O. Box 1738, 3000 DR Rotterdam, the Netherlands
Fax: +31-10-4089162
E-mail: andygykz97@gmail.com
E-mail: rinacheng1999@gmail.com
E-mail: frasincar@ese.eur.nl
*Corresponding author

Abstract: Nowadays, the Web is an essential hub for gathering comments on entities and their associated aspects. In this paper we propose a model which is capable of extracting these opinions and predicting the sentiment scores in aspect-level sentiment mining. In our two-step approach, a lexicalized domain ontology is firstly applied for sentiment classification. If the result is inconclusive from the first step, the backup model Double Rotatory Attention Mechanism is applied, which utilizes deep contextual word embeddings to better capture the (multi-)word semantics in the given text. This study contributes to the current research by introducing novel repetition and rotatory structures to refine the attention mechanism. It is shown that our model outperforms state-of-the-art methods on the datasets of SemEval 2015 and SemEval 2016.

Keywords: LCR-Rot; Double Rotatory Attention; Contextual Word Embeddings; Sentiment Analysis

Biographical notes: Guangyao Zhou obtained his B.Sc degree in econometrics and operations research from Erasmus University Rotterdam, Netherlands, in 2021. Since 2021, he is a research assistant at Chinese Academy of Sciences. His research interests are in econometrics and machine learning.

Jingyi Cheng obtained her B.Sc degree in econometrics and operations research from Erasmus University Rotterdam, Netherlands, in 2021. Her research interests are related to decision modeling in artificial intelligence, complex system, and network science. She is a current graduate student in computational science at University of Amsterdam.

Flavius Frasincar obtained the M.Sc. degree in computer science from Politehnica University Bucharest, Romania, in 1998. In 2000, he received the professional doctorate degree in software engineering from Eindhoven University of Technology, the Netherlands. He got the Ph.D. degree in computer science from Eindhoven University of Technology, the Netherlands, in 2005. From 2005, he is an Assistant Professor in information systems at Erasmus University Rotterdam, the Netherlands. He has published in numerous conferences and journals in the areas of databases, Web information

systems, personalization, machine learning, computational linguistics, and the Semantic Web. He is a member of the editorial boards of Decision Support Systems, Information Processing & Management, International Journal of Web Engineering and Technology, and Computational Linguistics in the Netherlands Journal, and co-editor-in-chief of the Journal of Web Engineering.

1 Introduction

On the social Web, people are encouraged to actively share their opinions with the rest of the world conveniently. Consequently, the amount of review data grows explosively, and the process of extracting valuable information has become challenging (Schwartz and Ward, 2004). To solve this, sentiment analysis can be applied to process data automatically. As an important part of opinion mining, sentiment analysis aims to distill the underlying sentiments in textual information, and to present results comprehensively and coherently (Liu, 2012).

While there are many different levels of sentiment analysis, this paper focuses on aspect-based sentiment analysis (ABSA). Instead of giving a general sentiment score for a piece of text, ABSA aims to predict the sentiment scores towards the captured aspects of the discussed entity. Hence, ABSA helps to generate more information from the given text (Schouten and Frasincar, 2015). With the implementation of ABSA, it becomes easier for companies to know about consumers' opinions towards specific aspects of their service or products, which helps to increase business values via making use of customer feedback effectively. Likewise, potential customers can benefit from fine-grained reviews before making a purchase (Liu, 2012).

A classic ABSA approach includes three main tasks, namely, target extraction, aspect detection, and sentiment classification. While the target extraction identifies the attributes of the addressed entity, aspect detection finds the relevant aspects of the focused entity. For instance, '*service*' is an aspect word represented in the review "*The service does not live up to the expectation from a Michelin-starred restaurant*". Next, a sentiment score is assigned to the aspect found in the last step via sentiment analysis. In this review example, the sentiment of the related aspect '*service*' is negative.

The objective here is the choice of techniques that can be used to perform ABSA accurately and efficiently. Summarized by Schouten and Frasincar (2015), there are mainly three types of methods, i.e., knowledge-based methods, machine learning approaches, and hybrid methods combining the former two. Purely machine learning methods have been successful in some domains (Liu et al., 2018; Maas et al., 2011; Tang et al., 2014), and knowledge-based models are proved to be useful as well (Schuller and Knaup, 2011). Nonetheless, these methods have their downsides. Whereas knowledge-based methods utilize language and domain features, such as frequency, syntax, and ontology, they require huge manual labor on constructing a thorough knowledge repository. While machine learning methods gain flexibility, robustness, and help to avoid manual work, they often need a large amount of training data to perform well, which is not ideal for targeted entities or aspects with less available information. Recently, deep learning methods also show potential for providing state-of-the-art results to Natural Language Processing (NLP) classification tasks (LeCun et al., 2015). While rule-based classifiers require more human labor for domain knowledge input, deep learning gains flexibility from its end-to-end training (Li, 2017). Zheng and Xia (2018) propose a left-center-right separated neural network with rotatory attention (LCR-Rot), which contains three LSTMs matching the left context, target phrase, and right context of a review. The rotatory attention mechanism is designed to model the relation between the parts. It utilizes target2context attention to identify the most indicative sentiment words in left/right contexts, then identifies the most relevant word in the target phrase using context2target attention (Zheng and Xia, 2018).

The remaining question is how to design a hybrid approach to achieve better general performance in both efficiency and accuracy for ABSA. An example of a successful two-step sentiment classification method, as described by Schouten et al. (2017) and Schouten and Frasincar (2018), combines knowledge-based and Support Vector Machine (SVM) techniques. In these models, domain knowledge is modeled into an ontology, which is firstly used to capture the sentiment of a target. The SVM is triggered when the polarity of a target cannot be concluded from the first step. By replacing the backup model with the neural attention model, Meškelė and Frasincar (2019) improve the model performance. Wallaart and Frasincar (2019) introduce the Hybrid Approach for Aspect Based Sentiment Analysis (HAABSA) approach using LCR-Rot as a backup solution to handle unclassified instances by a lexicalized domain ontology. Furthermore, Wallaart and Frasincar (2019) propose two additional attention designs to improve LCR-Rot. The first design called Inversed LCR-Rot inverts the order of the original rotatory attention mechanism by applying context2target before target2context attention algorithm. The other design called Multi-Hop LCR-Rot intends to repeat the rotatory attention mechanism multiple times. So the output from the context2target algorithm is utilized as the input for the target2context algorithm in the next repetition. According to the experiments by Wallaart and Frasincar (2019) using SemEval 2015 and SemEval 2016 dataset, the hybrid approach using a domain ontology with Multi-Hop LCR-Rot as backup solution performs the best. The study of Wallaart and Frasincar (2019) is then extended by Truşcǎ et al. (2020), in which the original non-contextual word embeddings are substituted with deep contextual word embeddings, and another hierarchical attention layer is added at the end of Multi-Hop LCR-Rot, called HAABSA++. It improves the HAABSA accuracy for both SemEval 2015 and SemEval 2016 datasets.

The implementation of repeated attention has shown to be successful for both Multi-Hop LCR-Rot (Wallaart and Frasincar, 2019), and Multi-Hop LCR-Rot with hierarchical attention structure (Truşcǎ et al., 2020). At the current LCR-Rot approach, it always starts with target2context operation by taking target information as input, which is then followed by the context2target process with the context attention outcomes from the previous step as inputs. And this two-step procedure is sequentially repeated several times in a Multi-Hop LCR-Rot model. Yet, the possibility of performing target2context and context2target processes simultaneously with both target and context information has not been discussed before. This study, therefore, sets out to investigate this possibility and evaluate the effect on the accuracy of such structure.

Built upon the work by Zheng and Xia (2018), Wallaart and Frasincar (2019), and Truşcǎ et al. (2020), we propose a hybrid approach for aspect-based sentiment analysis, where a lexicalized domain ontology is firstly employed with Multi-Hop LCR-Rot with double rotatory attention as the backup algorithm. Our research seeks to refine the backup mechanism LCR-Rot based structure by improving the utilization of information and training efficiency with our proposed attention mechanisms. We propose two novel structures in this study. By Double Rotatory Attention Mechanism, the target, as well as context information, are taken as input simultaneously to each iteration of rotatory attention process. In this design, we perform two times rotatory attention iterations, after which the outcomes are refined by the Attention Mechanism. Next, an extra step of attention rotation is added to refine the representation outcomes from the previous structure. We also propose Two Single Rotatory Structure as the final Rotatory Attention Mechanism. Inspired by HAABSA++ (Truşcă et al., 2020), a three-step rotatory with a hierarchical attention structure is considered as an alternative here. Our study provides new insights into performing multi-hop mechanism on the LCR-Rot model in an information-efficient manner. Meanwhile, the success achieved by our proposed attention structures based on the example of the LCR-Rot model can shed the light on the future study of other types of attention-based neural networks.

The remainder of this paper is structured as follows. In Section 2, we discuss the related literature with respect to sentiment analysis utilizing domain ontology and neural attention models. Then, in Section 3, a description of the proposed methodology and extensions is present. Next, we provide an overview of the used datasets of this paper in Section 4. In Section 5, we present the experiment outcomes and the comparative evaluation of different models. Lastly, Section 6 concludes this paper and provides directional suggestions for future research.

2 Related Literature

As the amount of online reviews increases, research over feature-based analysis is becoming increasingly important to the industry. Introduced by Hu and Liu (2004), a feature-based summary can be decomposed into three subtasks: (1) extracting the described features for any given review (Chen et al., 2014; Poria et al., 2016); (2) identifying the polarity values for each feature of review sentences (Das and Chen, 2007); (3) generating a summary over all the extracted information. The last subtask is also known as Aspect-Based Sentiment Analysis (ABSA) (Pontiki et al., 2015, 2016). ABSA takes both sentiment and its target information into consideration for classification. In the sample review "The food is delicious but the service is not good enough.", the sentiment for food is positive, while the sentiment for service is negative. Readers can easily perform the sentiment matching by correlating an aspect to its descriptions, using grammar knowledge. But for models, the semantic relatedness of a target with its surrounding context information needs to be determined (Zhang et al., 2018). The main solutions for ABSA are classified into knowledge-based methods, machine learning methods, and hybrid methods that utilize the first two (Schouten and Frasincar, 2015).

To improve the performance of machine learning methods, approaches, such as the Bag-of-Words (BoW) and ontology reasoning (Schouten et al., 2017), are developed to simplify the feature detection process. BoW translates input reviews into binary vectors that represent a given word's appearance in selected sentences. To exploit the common domain knowledge information, one can also use a domain ontology for sentiment classification, which can be created manually (Schouten and Frasincar, 2018; Schouten et al., 2017), semi-automatically (Coden et al., 2014; Zhuang et al., 2020), or automatically (Alani et al., 2003). Ontology has achieved success in the information retrieval tasks of different domains. Li et al. (2008) introduce an engineering ontology-based algorithm to retrieve unstructured engineering documents or drawings based on the associated textual information. Zhang et al. (2009) propose an ontology-based system to match the product information to customers for E-commerce businesses. In the cases where multiple topics are present, such as sentiment analysis for tweets (Kontopoulos et al., 2013), ontology-based methods can match domain-specific sentiment scores to each domain concept appearing in the text. In the field of text mining, the ontology-based approach is also shown useful. OTTO (OnTology-based Text mining framewOrk) proposed by Bloehdorn et al. (2005) applies ontology learning techniques (Maedche and Staab, 2004) to construct a target ontology for supervised or unsupervised text categorization. Applying ontology reasoning sequentially with BoW or other regular machine learning methods has shown to be useful in enhancing accuracy for the ABSA tasks (Schouten and Frasincar, 2018).

In order to better capture semantics, neural attention models have gained popularity in recent research (Zhang et al., 2018). Given any text input, it is a natural next step to consider how to allocate neural attention effectively. To make use of the target information, Target-Dependent LSTM (TD-LSTM) and Target-Connection LSTM (TC-LSTM) are proposed by Tang et al. (2015). Utilizing a context attention mechanism, Liu et al. (2018) propose the Content Attention Based Aspect-based Sentiment Classification (CABASC) model, which classifies based on both the order of words and their intra-correlations. The possibility of separating attention over sentences into parts is also explored. A Left-Center-Right separated neural network with Rotatory attention (LCR-Rot) is proposed in Zheng and Xia (2018), where the interactions among aspect (center) and left/right contexts are strengthened to identify aspects and their most relevant context. Since hierarchical models have gained success in representation learning, a hierarchical bidirectional LSTM model is introduced by Ruder et al. (2016) for review-level analysis. It can leverage both intra- and inter-sentence relationships while handling a large number of preceding and successive sentences. Word embeddings results are often used as input features for deep learning methods (Collobert et al., 2011). As a technique for feature learning, word embeddings transform words into continuous real numbers vectors. Neural networks are commonly used for word embeddings training (Bengio et al., 2003; Mikolov, Sutskever, Chen, Corrado and Dean, 2013; Mikolov, Chen, Corrado and Dean, 2013; Mnih et al., 2013; Morin and Bengio, 2005). Alternatively, matrix factorization can also be used for word embeddings training (Huang et al., 2012; Pennington et al., 2014).

The attention mechanism has become an important piece of recent neural architectures. Rather than processing the whole input text into fix-length representation vectors like traditional Recurrent Neural Network (RNN) and LSTMs, the attention mechanism enables the model to learn the more important parts on the base of the input text and its memory. An Interactive Attention Network (IAN) is proposed by Ma et al. (2017), which interactively utilizes two attention networks for the detection of word importance. Furthermore, Attention-over-Attention (AOA) neural models are employed to simplify IAN and avoid the iteration-oblivious problems caused by the pooling operations in IAN (Huang et al., 2018; Cui et al., 2016). Devlin et al. (2019) propose the contextual word embeddings Bidirectional Encoder Representations from Transformers (BERT), which allows model fine-tuning. Since BERT adopts the self-attention mechanism to unify two-stage relationships within text pairs, fine-tuning BERT is a logical next step to encode the bidirectional cross-attention between two sentences.

To obtain a similar level of accuracy, fine-tuning BERT is considered a less expensive solution than pre-training models.

Hybrid models that sequentially combine the knowledge-based approach and the machine learning approach have been proven to be highly effective (Schouten et al., 2017). In (Schouten and Frasincar, 2018), the proposed two-step sentiment classifier employs a domain-based ontology for preliminary classification. And the BoW trained SVM model is applied as a backup solution when the ontology is inconclusive. This backup solution is enhanced by Wallaart and Frasincar (2019) using Left-Center-Right separated neural network with Rotatory attention (LCR-Rot). The latter approach is further improved with deep contextual word embeddings from BERT and hierarchical attention, as introduced by Truşcă et al. (2020).

3 Methodology

In this paper, we propose a *Hybrid Approach for Aspect-Based Sentiment Analysis using Double Rotatory Attention* (HAABSA-DRA) as our two-step hybrid method for aspect-based sentiment classification. Firstly, a domain sentiment ontology is applied to predict the polarity scores of target reviews. If the first step is unsuccessful, a neural network model is used as a backup solution. In Section 3.1, we explain the ontology structure and how is the ontology developed for sentiment analysis. In Section 3.2, we give a short description about the employed word embeddings. In Sections 3.3, 3.4, and 3.5, we present the employed rotatory attention mechanism, extensions, and loss function, respectively.

3.1 Ontology-Based Rules

The lexicalised domain ontology makes use of predefined classes, their interrelations, and axioms to predict the sentiment score of given aspects. Inherited from Schouten et al. (2017), the adopted ontology of this paper is a manually constructed domain specification for sentiment polarities of aspects, which is seen to be more reliable for the correctness of specified elements than the ones created semi-automatically or automatically. Three main classes are involved in this ontology: the *SentimentValue* consists of *Positive* and *Negative* as its subclasses, so *Neutral* is not included due to its ambiguous nature; the *AspectMention* identifies aspects mentioned; the *SentimentMention* matches sentiment expressions to different categories. Three ontology-based types of rules introduced by Schouten and Frasincar (2018) are used to compute the sentiment value of each aspect and are elaborated on below.

The first rule type returns the generic sentiments of the target aspect concerning its sentiment expression, e.g., ‘unacceptable’ goes to the class *Negative*. If the first rule type does not apply, the second rule type is used to identify aspect-specific sentiment expressions, where sentiment is assigned only if the aspect and the corresponding expression belong to the same aspect-category, e.g., ‘crowded’ is *Negative* under the scope of *AmbienceGeneral* but it is not defined for the aspect *FoodPrice*. Lastly, the third rule type is designed to detect the expression with a varying context-dependent sentiment. It returns sentiment based on the aspect-sentiment category such expression falls into (e.g., *WarmCola* is *Negative*, while *WarmCoffee* is *Positive*). Those three rule types are ordered and exclusive, namely, a higher-order rule type can only be applied if all the previous rule types are not suitable. In other words, a given sentiment will go through the third rule type only when the first and second rule types fail to detect sentiment. For the case where the negation word is presented within a range of lexical distance (3 words apart), the sentiment polarity value is flipped to the opposite (Yu et al., 2011).

However, the introduced ontology-based rule is inconclusive in the following two cases: (1) Fail-to-catch, which is caused by limit coverage (2) Conflicting classification, where both *Positive* and *Negative* are predicted for the target aspect. Next, we can describe the various classes with a specific property attached, such as ENVIRONMENT#BAD. The property is usually an adjective, and adjectives contain emotions, such as commendatory, derogatory or neutral. If it is a commendatory word, then we can judge that the reviewer's emotion is positive; if it is a derogatory word, the reviewer is not satisfied in this aspect, and his emotion is negative. Finally, according to the nature of adjectives, we label the classes as positive, negative, or neutral. Let us give an example to illustrate the workings of our ontology. The review is “the waiters are very enthusiastic”, because “waiters” are a subclass of “services”, while “enthusiastic” is a positive word. Therefore, “services" is labeled positive.

Most classes have lexical representations attached that allow for easy concept discovery in text. The restaurant domain ontology is created for our experiments following the method introduced by Schouten and Frasincar (2018).

3.2 Word Embeddings

Word embeddings are dense and low-dimensional vector representations of words that capture the similarity of words with close meanings. One of the first neural network-based word embedding methods was introduced by Google in 2013 (Mikolov, Sutskever, Chen, Corrado and Dean, 2013), which was successfully applied to various NLP tasks. Word embeddings are generally categorized into two types, non-contextual word embeddings, and contextual word embeddings. Non-contextual word embeddings, such as $word2vec$ and $GloVe$, consider that each word is unique and has a similar meaning regardless of the context. On the other hand, contextual word embeddings consider the context of words together with the words themselves, such that they better summarize the semantic information from the text. For example, contextual word embeddings give the word ‘*apple*’ different vector representations for ‘*I like apple pie*’ and ‘*I own an Apple MacBook.*’. The most used contextual word embeddings are $ELMo$ and $BERT$ (Google's Bidirectional Encoder Representations from Transformer (Devlin et al., 2019)). Since $BERT$ achieves state-of-the-art results in recent researches, we adopt BERT as our word embedding method in this paper.

As introduced by Devlin et al. (2019), the $BERT$ model is pre-trained with two unsupervised tasks with BookCorpus and Wikipedia dumps: Masked Language Model (MLM) and Next Sentence Prediction (NSP). By masking part of the input tokens randomly, MLM trains a deep bidirectional transformer to predict the masked words, and NSP is trained to understand sentence relationships, which has significant usage in Question Answering and Natural Language Inference applications. Input presentations are generated with WordPiece embeddings (Wu et al., 2016). The beginning of each input sentence is masked with a classification token $[CLS]$, and separator $[SEP]$ tokens are inserted to split the input sequence into sentences. Word embedding input vectors are generated by averaging three vectors: token embeddings (unique for each word), segment embeddings (indicating which sentence the target word corresponds to), and position embeddings (the location of the word in this input sequence). Given that L denotes the number of Transformer blocks, H denotes the hidden states, and A stands for the amount of self-attention heads, we employ the Base BERT model with L = 12, H = 768, A = 12 (Devlin et al., 2019). Since the summation of the last four layers is shown to give promising results (Devlin et al., 2019), the final representation of Word$_i$ is:

$$BERT_i = \sum_{j=9}^{12} H_{i,j} \tag{1}$$

3.3 Left-Center-Right Separated Neural Network with Rotatory Attention Attention Mechanism

The LCR-Rot model by Zheng and Xia (2018) makes use of the left-center-right structure to detect the relationship between semantic expressions and their contexts. Extending basic structure of LCR-Rot, Wallaart and Frasincar (2019) introduce a Multi-Hop LCR-Rot classifier by refining the model with repeated attention, a new classifier that obtains a better performance. As an extension, Truşcă et al. (2020) introduce a hybrid approach with BERT contented embeddings and hierarchical attention layers based on Multi-Hop LCR-Rot and achieves state-of-the-art results on SemEval 2015 and SemEval 2016. In (Huang et al., 2018) and (Cui et al., 2016), attention-over-attention (AOA) captures the interaction of aspects and sentences. And it automatically allocates more attention to the important parts in sentences. In this paper, we propose a *Left-Center-Right Separated Neural Network with Rotatory Attention and Attention* (LCR-DRA). It is adapted from the LCR-Rot algorithm using bi-directional hidden states of left/right context and target phrase simultaneously as input and enhanced by a structure with three rotatory steps and an attention operation.

Section 3.3.1 gives an introduction to the initial *Left-Center-Right Separated Neural Network* model. Since the Multi-Hop structure of Wallaart and Frasincar (2019) achieves good results, this paper inherits the idea of 'repeat for enhancement' to retrieve improved attention weights in an iterative nature. The model proposed by us further exploits the rotatory attention mechanism, which is repeated sequentially three times. The newly-designed rotatory structure *Double Rotatory Attention* Mechanism is presented in Section 3.3.2, where an attention mechanism is introduced to boost representations.

3.3.1 Left-Center-Right Separated Neural Network

To start with, we define a sentence with N words as S = $[s1, s2, ..., s_N]$. Each sentence can be roughly considered as a lexical combination of three parts, where the review subject is estimated to be located at the target phrase and the descriptions of the subject at the left/right contexts. So the essential next step is to detect the most important target-context pairs of sentences, especially for the cases that the basic ontology-based model fails to detect.

The LCR model splits the sentence into three parts: left context $[s_l^1, s_l^2, ..., s_l^L]$, target phrase $[s_t^1, s_t^2, ..., s_t^T]$, and right context $[s_r^1, s_r^2, ..., s_r^R]$, where L, T, R are the amount of words contained in each part and the sum of L, T, and R is equal to N. Bi-directional long-short-term-memory (Bi-LSTM) is utilized at the next step to generate left-, target-, and right-hidden layers representations. The LSTM networks are specialized in long-term memory, and the bi-directional structure stores the contextual information for both directions. Notably, the input of Bi-LSTM is d-dimensional word embeddings, with which Bi-LSTM gives the bi-directional hidden states $\underset{2d\times L}{[h_l^1, h_l^2, ..., h_l^L]}$, $\underset{2d\times T}{[h_t^1, h_t^2, ..., h_t^T]}$ and $\underset{2d\times R}{[h_r^1, h_r^2, ..., h_r^R]}$, respectively, for left/right context and target phrase.

3.3.2 Double Rotatory Attention Mechanism

The *Double Rotatory Attention (DRA)* Mechanism is designed to capture the most indicative word and its representation in the target phase and left/right contexts by applying two attention rotations sequentially, where higher attention is rewarded to the correct target-context pairs during the supervised training process. Nevertheless, an attention mechanism is adopted to refine the attention scores to emphasize the significance of the most contributing parts of the reviews. An illustration of the architecture of this model is provided in Figure 1. For clarification, the pseudocode of this part is elaborated at Algorithm 1.

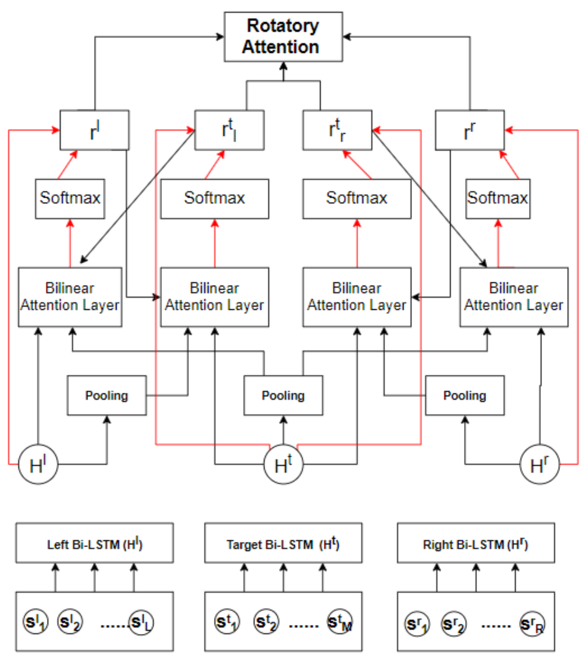

Figure 1: Double Rotatory Attention Mechanism

To achieve the learning goal, we can start with applying the first two-step rotatory attention mechanism over the three parts of hidden states, where the type of this layer is average pooling:

$$r_p^t = pooling([\underset{2d\times 1}{h_1^t}, \underset{2d\times 1}{h_2^t}, ..., \underset{2d\times 1}{h_T^t}]); \tag{2}$$

$$r_p^l = pooling([\underset{2d\times1}{h_1^l}, \underset{2d\times1}{h_2^l}, ..., \underset{2d\times1}{h_L^l}]); \tag{3}$$

$$r_p^r = pooling([\underset{2d\times1}{h_1^r}, \underset{2d\times1}{h_2^r}, ..., \underset{2d\times1}{h_R^r}]). \tag{4}$$

First, to detect the most indicative words in contexts, we utilize the context information to obtain better target representations. This step is named *Context2Target Attention* mechanism. Take the example of the left context representation, an attention function f is defined as:

$$\underset{1\times1}{f(h_i^t, r_p^l)} = \tanh(\underset{1\times2d}{h_i^{t'}} \times \underset{2d\times2d}{W_t^l} \times \underset{2d\times1}{r_p^l} + \underset{1\times1}{b_t^l}). \tag{5}$$

And for *Target2Context Attention* mechanism, the attention function f for left target-aware representation is defined as:

$$\underset{1\times1}{f(h_i^l, r_p^t)} = \tanh(\underset{1\times2d}{h_i^{l'}} \times \underset{2d\times2d}{W_c^l} \times \underset{2d\times1}{r_p^t} + \underset{1\times1}{b_c^l}), \tag{6}$$

where W_t^l (W_c^l) is a weight matrix, r_p^l (r_p^t) is a left-context representation (target representation) initialized by average pooling operation, b_t^l (b_c^l) is the bias term, and $2d$ represents the dimension of the i^{th} hidden state h_i^t of target phrase for $i = 1, ..., T$. *tanh* is adopted as an activation function, which is commonly used to help converge faster in LSTM models and let gradient computation be less expensive, as compared to other basic activation function options (Vijayaprabakaran and Sathiyamurthy, 2020).

Given the hidden left and right context representation $h^l \in \mathbb{R}^{2d\times L}$, $h^r \in \mathbb{R}^{2d\times R}$ and hidden target representation $h^t \in \mathbb{R}^{2d\times T}$, the target attention scores $f(h_j^t, r_p^l)$ are fed into a softmax function to obtain the attention normalised scores $a_i^{t_l}$,

$$a_i^{t_l} = \frac{\exp(f(h_j^t, r_p^l))}{\sum_{j=1}^{L} \exp(f(h_j^t, r_p^l))}. \tag{7}$$

Likewise, attention scores a_i^l and a_i^r are obtained with softmax operation for the first time,

$$a_i^l = \frac{\exp(f(h_i^l, r_p^t))}{\sum_j \exp(f(h_j^l, r_p^t))}. \tag{8}$$

Lastly, the left-aware target representation r_l^t is retrieved as the sum of the word hidden states scaled by attention scores in the target phrase,

$$\underset{2d\times1}{r_l^t} = \sum_{i=1}^{M} \underset{1\times1}{a_i^{t_l}} \times \underset{2d\times1}{h_i^t}. \tag{9}$$

Similarly, left context representation is retrieved by:

$$\underset{2d\times1}{r^l} = \sum_{i=1}^{L} \underset{1\times1}{a_i^l} \times \underset{2d\times1}{h_i^l}. \tag{10}$$

By repeating the equations (5)-(10) in the same fashion, we can get the right-aware target representation r_r^t and right-aware context representation r^r.

The second part of the double rotatory attention is applied by repeating the rotatory attention mechanism analogous to equation (5)-(10): r_p^l in equation (5), (7) is replaced by updated r^l from the primary rotatory attention process; likewise, r_p^t in equation (6), (8) is replaced by r_l^t; equations (9)-(10) remain unchanged. Similar updates are done for the right context.

input : Bi-directional hidden states H_l, H_t, H_r
output: Context/target representations r^l, r^r, r_r^t, r_l^t

begin

/* **Pooling** */
Generate r_p^l, based on *pooling*(H_l);
Generate r_p^t, based on *pooling*(H_t);
Generate r_p^r, based on *pooling*(H_p);
$r_{(l)}^{t_p} \leftarrow r_p^t$; $r_{(r)}^{t_p} \leftarrow r_p^t$;

iterate for 2 cycles

/* **Binary Attention Layer** */
Generate $f(h_i^t, r_p^l)$, based on *Context2Tagret*$(h_i^t, r_p^l, W_t^l, b_t^l)$;
Generate $f(h_i^t, r_p^r)$, based on *Context2Tagret*$(h_i^t, r_p^r, W_t^l, b_t^r)$;
Generate $f(h_i^l, r_{(l)}^{t_p})$, based on *Target2Context*$(h_i^l, r_{(l)}^{t_p}, W_c^l, b_c^l)$;
Generate $f(h_i^r, r_{(r)}^{t_p})$, based on *Target2Context*$(h_i^r, r_{(r)}^{t_p}, W_c^r, b_c^r)$;

/* **Softmax** */
Generate $a_i^{t_l}$, based on *Softmax*$(f(h_i^t, r_p^l))$;
Generate $a_i^{t_r}$, based on *Softmax*$(f(h_i^t, r_p^r))$;
Generate a_i^l, based on *Softmax*$(f(h_i^l, r_{(l)}^{t_p}))$;
Generate a_i^r, based on *Softmax*$(f(h_i^r, r_{(r)}^{t_p}))$;

/* **Context/Target Representation** */
Compute left-aware target representation r_l^t, based on $\sum_{i=1}^{M} a_i^{t_l} \times h_i^t$;
Compute right-aware target representation r_r^t, based on $\sum_{i=1}^{M} a_i^{t_r} \times h_i^t$;
Compute left-aware context representation r^l, based on $\sum_{i=1}^{L} a_i^l \times h_i^l$;
Compute right-aware context representation r^r, based on $\sum_{i=1}^{L} a_i^r \times h_i^r$;

/* **Reassignment** */
$r_p^l \leftarrow r^l$; $r_p^r \leftarrow r^r$; $r_{(l)}^{t_p} \leftarrow r_l^t$; $r_{(r)}^{t_p} \leftarrow r_r^t$;

end

Denote current representation $v^i \in \{r_l^t, r_r^t, r^l, r^r\}$;

end

Algorithm 1: Double Rotatory Attention Mechanism

3.4 Refining Rotatory Attention Mechanism

In this section, we propose to add a final layer of rotatory attention mechanism to generate the outcome, where the attention allocations within sentences and targets are further refined by using the representations obtained from *Double Rotatory Attention* Mechanism. The first type of rotatory attention mechanism employs a similar structure as is already introduced by HAABSA++ (Truşcă et al., 2020). That is, it consists of a single LCR-Rot with Hierarchical Attention. However, the drawback of this structure is that only half of what the *Double Rotatory Attention* Mechanism generates can be used, which causes a waste of information. In order to fully utilize outputs from the last step, we propose the *Two Single Rotatory* structure that uses two rotatory structures to efficiently and separately process the output from the *Double Rotatory Attention* Mechanism. The first and second type of rotatory attention mechanism are referred as Rotatory Attention Mechanism Type I and II respectively in the following.

3.4.1 Rotatory Attention Mechanism Type I: Three-step Rotatory with Hierarchical Attention Structure

The first type rotatory attention mechanism performs in three steps, where Target2Context Attention Mechanism is the first step, Context2Target Attention Mechanism is the second step and hierarchical attention is employed at the third step. In the first step, a new pair of context-aware representations r^l, r^r are jointly generated by target-aware representations r^{t_l}, r^{t_r} obtained from *Double Rotatory Attention* Mechanism and hidden left/right context representations H^l, H^r. Then, the newly generated r^l, r^r are once again combined with target representation H^t to create a pair of refined target-aware representations r^t_l, r^t_r.

The steps named Target2Context Attention Mechanism are updated with equations (11), (12) and (13),

$$\underset{1\times 1}{f(h^l_i, r^t_l)} = \tanh(\underset{1\times 2d}{h^{l'}_i} \times \underset{2d\times 2d}{W^l_c} \times \underset{2d\times 1}{r^t_l} + \underset{1\times 1}{b^l_c}); \tag{11}$$

$$a^l_i = \frac{\exp(f(h^l_j, r^t_l))}{\sum_{j=1}^{L} \exp(f(h^l_j, r^t_l))}; \tag{12}$$

$$\underset{2d\times 1}{r^l} = \sum_{i=1}^{L} \underset{1\times 1}{a^l_i} \times \underset{2d\times 1}{h^l_i}. \tag{13}$$

The second step is Context2Target Attention Mechanism, which uses equation (14), (15) and (16),

$$\underset{1\times 1}{f(h^t_i, r^l)} = \tanh(\underset{1\times 2d}{h^{t'}_i} \times \underset{2d\times 2d}{W^l_t} \times \underset{2d\times 1}{r^l} + \underset{1\times 1}{b^l_t}); \tag{14}$$

$$a^{t_l}_i = \frac{\exp(f(h^t_j, r^l))}{\sum_{j=1}^{L} \exp(f(h^t_j, r^l))}; \tag{15}$$

$$\underset{2d\times 1}{r_l^t} = \sum_{i=1}^{M} \underset{1\times 1}{a_i^{t_l}} \times \underset{2d\times 1}{h_i^t}. \tag{16}$$

One disadvantage of the proposed model in Section 3.3 is that it only adopts local information within groups Left, Target, and Right to compute the Context2Target and Target2Context vectors. To integrate the information, hierarchical attention is introduced to provide an aggregated representation of the input sentences, which are used to compute each context2target and target2context with a relevance score at the sentence level.

First, an attention function f is given as:

$$\underset{1\times 1}{f(\phi^i)} = \tanh(\underset{1\times 2d}{\phi^{i'}} \times \underset{2d\times 1}{W} + \underset{1\times 1}{\delta}), \tag{17}$$

where $\phi^i \in \{r^r, r^l, r_r^t, r_l^t\}$ is representation vectors generated from the first two steps, W is a weight vector, and δ is the bias.

Then, the attention scores α_i is computed with a softmax function using the result of attention function f for each input ϕ_i:

$$\alpha^i = \frac{\exp(f(\phi^i))}{\sum_{j=1}^{4} \exp(f(\phi^j))}. \tag{18}$$

Finally, the target context2target or target2context vector is updated with attention scores from last step:

$$\underset{2d\times 1}{\phi^i} = \underset{1\times 1}{\alpha^i} \times \underset{2d\times 1}{\phi^i}. \tag{19}$$

By repeating the attention weighting for the final four vectors of the rotatory attention, hierarchical attention is added to refine the model as the last step (Truşcă et al., 2020). The architecture of hierarchical attention is illustrated in Figure 2. The use of hierarchical attention is to compensate for the inefficiency of information waste caused by not inheriting context representation r^l and r^r as inputs. The pseudocode of the Rotatory Attention Mechanism Type I is presented in Algorithm 2.

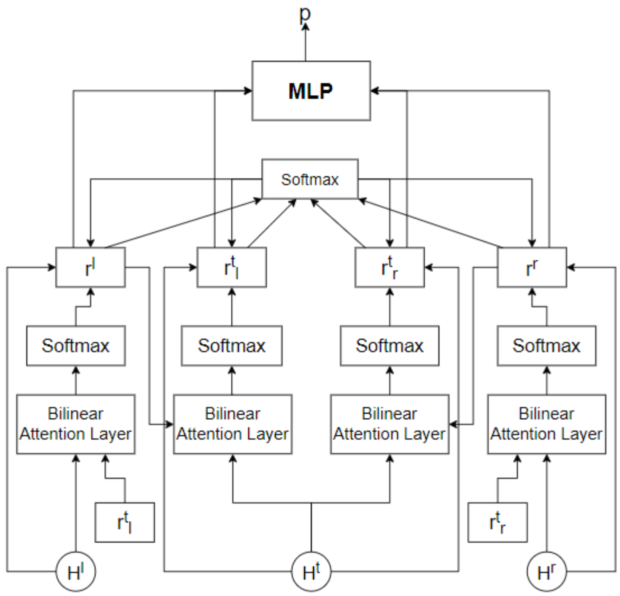

Figure 2: Three-step Rotatory Attention with Hierarchical Attention

input : Target representation $v^i \in \{r^t_l, r^t_r\}$
output: Updated context/target representation $\phi^i \in \{r^t_l, r^t_r, r^l, r^r\}$

begin

/* **Target2Context Attention Mechanism** */

Update context representation r^l and r^r, based on *Target2Context Attention* Mechanism;

/* **Context2Target Attention Mechanism** */

Update target representation r^t_l and r^t_r, based on *Context2Target Attention* Mechanism;

Denote the re-updated representations as ϕ^i, where $\phi^i \in \{r^t_l, r^t_r, r^l, r^r\}$;

/* **Hierarchical Attention** */

for ϕ^i *in* $\{r^t_l, r^t_r, r^l, r^r\}$ **do**

Update $f(\phi^i)$ based on *sentiment polarity estimation*(ϕ^i, W, δ) ;

Update ϕ^i based on *Softmax*$(f(\phi^i))$ $\times \phi^i$;

end

end

Algorithm 2: Rotatory Attention Mechanism Type I: Three-step Rotatory with Hierarchical Attention Structure

3.4.2 Rotatory Attention Mechanism Type II: Two Single Rotatory Structures

One of the drawbacks of the *Three-step Rotatory with Hierarchical Attention Structure* is that only the target-aware representations r_l^t and r_r^t from the last segment are applied, which means a 'calculation waste' is caused by abandoning the context information. Thus, the contribution of the *Double Rotatory Attention* Mechanism is limited by the structure of the Type I Rotatory Attention Mechanism.

To utilize all four representations, an alternative solution is to use two single rotatory structures, shown in Figure 3. We propose that the left target/context-aware representations r_l^t, r_r^t, r^l, r^r are separately and simultaneously rotated with hidden target/context representations H_l, H_r, H_t, such that the final outcome benefits from taking all the target/context-aware information into account. In this case, the target/context-aware representations separately follow the exact equations (11)-(13)/(14)-(16) from Section 3.4.1. This alternative method is named Rotatory Attention Mechanism Type II. Its pseudocode is presented in Algorithm 3.

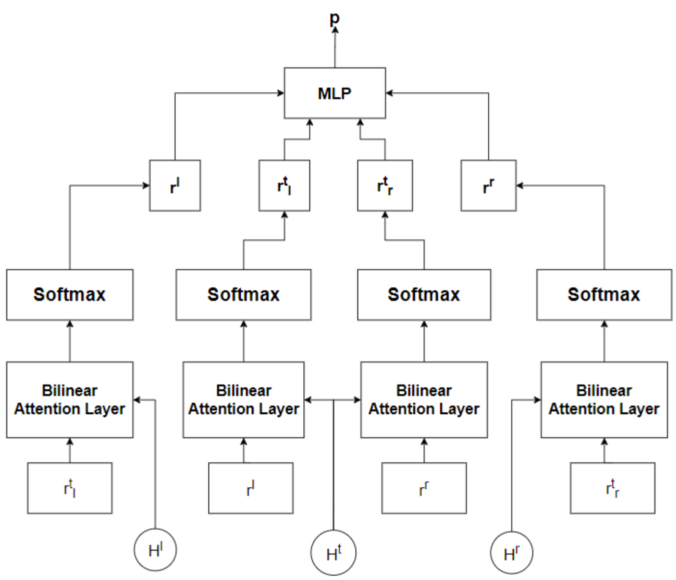

Figure 3: Rotatory Attention Mechanism Type II: Two Single Rotatory Structures

To show the full structure of our model and the benchmark model to exhibit the composition of the model and the differences between them, we exhibit the detail comparison in Figure 4 and Figure 5. Clearly, our redesign of the base model did not add overly complex structures, which means that the model complexity has not changed. Therefore, the training time of these models is almost the same.

3.5 Regularization and Loss Function

The supervised learning process is achieved by utilizing a backpropagation algorithm by minimizing a cross-entropy loss function with L2 regularization. Weight matrices and

input : Context/target representation $v^i \in \{r_l^t, r_r^t, r^l, r^r\}$
output: Updated context/target representation $\phi^i \in \{r_l^t, r_r^t, r^l, r^r\}$

begin
/* **Simultaneous Rotatory Attention Mechanism** */
do in parallel
Update context representation r^l and r^r, based on *Target2Context Attention* Mechanism;
Update target representation r_l^t and r_r^t, based on *Context2Target Attention* Mechanism;
end
Denote the re-updated representations as ϕ^i, where $\phi^i \in \{r_l^t, r_r^t, r^l, r^r\}$;
end

Algorithm 3: Rotatory Attention Mechanism Type II: Two Single Rotatory Structure

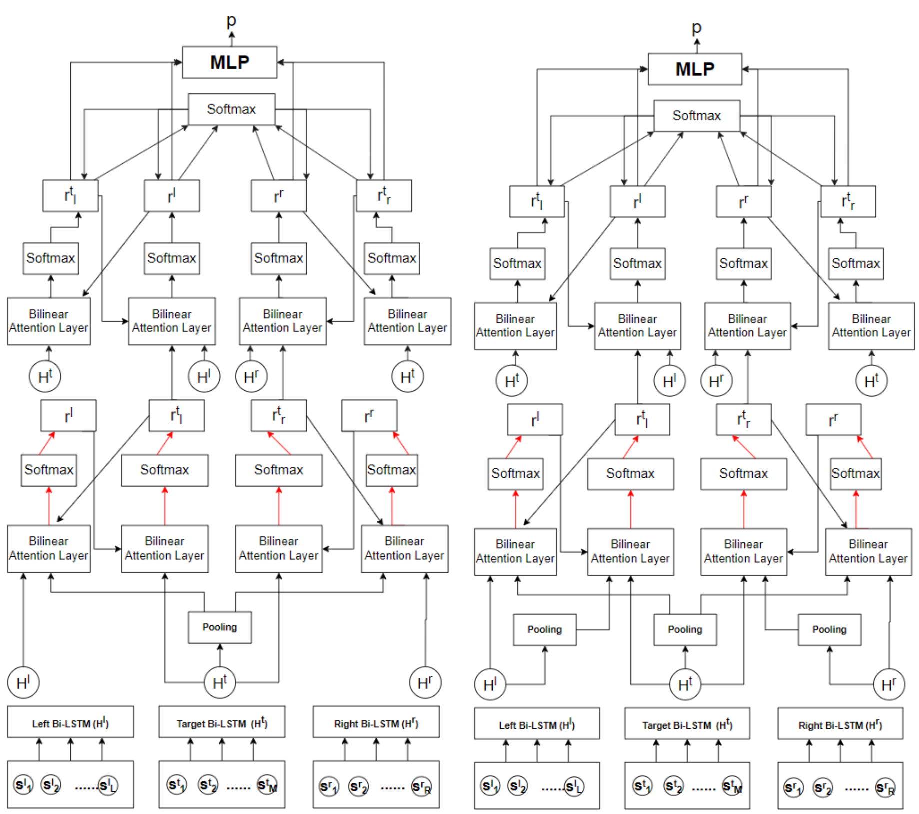

Figure 4: HAABSA++ (Truşcă et al., 2020) (Left) and HAABSA-DRA with Type I Attention (Right)

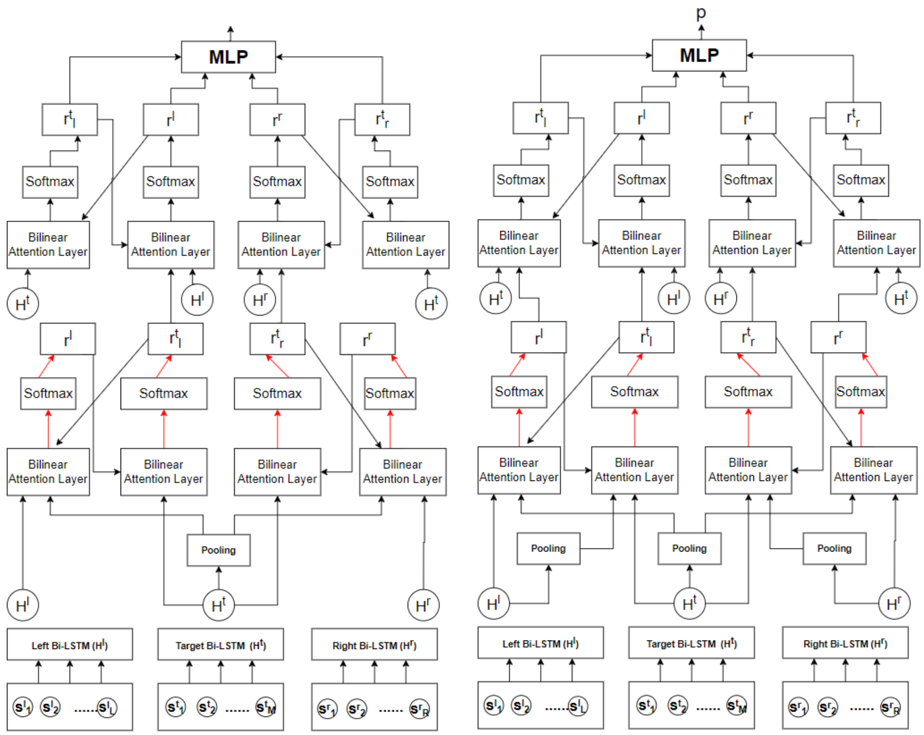

Figure 5: HAABSA (Wallaart and Frasincar, 2019) (Left) and HAABSA-DRA with Type II Attention (Right)

biases are updated by stochastic gradient descent with momentum through iterations, while they are first initialized with a uniform distribution. The hyperparameter learning rate, the momentum term, the dropout rate, and the L2-norm regularization term are required to be re-tuned to obtain the best performance of the model. In our study, a Tree-structured Parzen Estimator (TPE) is applied for the hyperparameter tuning process to get a fast convergence speed.

4 Datasets

For model training and providing comparable performance data, we select “Subtask 1 Restaurant Domain English Training Data” as training material and “Subtask 1 Restaurant Domain English Gold Annotations Data” as the test dataset from the SemEval 2016 Task 5 (Pontiki et al., 2016). Also, “2015 ABSA Restaurant Reviews - Train Data” and “2015 ABSA Restaurants Reviews - Test Data - Gold Annotations” from SemEval 2015 Task 12 (Pontiki et al., 2015) are included in this research to give a more general overview of model performance. Each review contains several sentences, and each sentence has one or more opinions. The associated targets of any given opinion are related to specific aspect categories. Such relationships determine the polarity values (positive, neutral, or negative) of aspects. Table 1 shows the polarity distribution of the datasets provided by SemEval 2015 and SemEval 2016. The aspect category distribution is shown at Figure 6. Obviously, the two datasets have similar distributions in the aspect category, but the sizes of the datasets are different. Besides, these datasets are imbalanced due to extreme differences in the polarity distribution.

Table 1 Sentiment Class Frequencies of Datasets

	Positive	**Neutral**	**Negative**
SemEval 2016 Train	70.2%	3.8%	26.0%
SemEval 2016 Test	74.3%	4.9%	20.8%
SemEval 2015 Train	72.4%	24.4%	3.2%
SemEval 2015 Test	53.7%	41.0%	5.3%

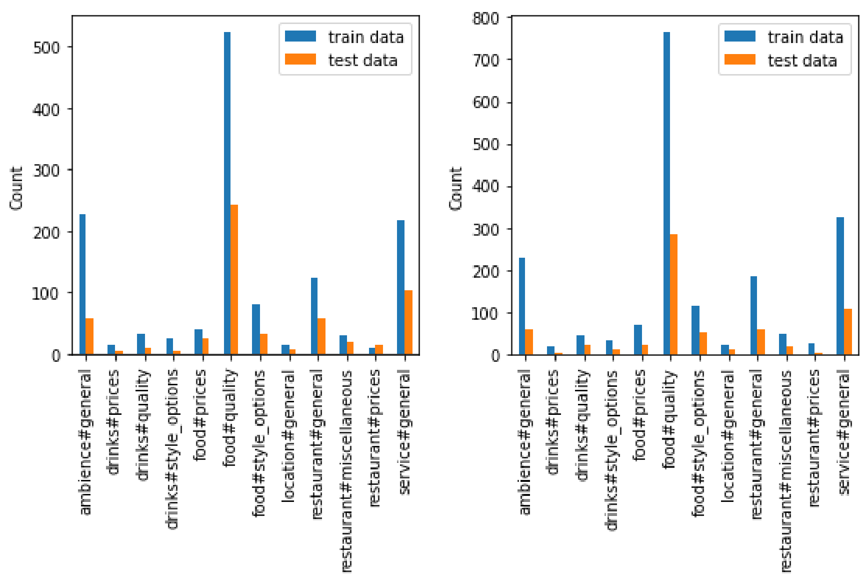

Figure 6: Aspect categories in SemEval 2015 (left panel) and SemEval 2016 (right panel)

5 Results

In this paper, we extend the baseline HAABSA model and evaluate it with two types of rotatory attention mechanism using both SemEval 2015 and SemEval 2016 datasets. Just like Truşcă et al. (2020), the first step of the hybrid model is the domain sentiment ontology to predict polarity; if unsuccessful, the second step is implemented using the proposed LCR-Rot neural network with extensions. To improve the accuracy, the BERT method of word embeddings is adopted, and hierarchical attention layers are added in Rotatory Attention Mechanism Type I, along with the *Left-Center-Right Separated Neural Network with Rotatory Attention and Attention Mechanism*. The whole process with Type I is similar to the hierarchical attention structure by HAABSA++ (Truşcă et al., 2020). More precisely, we use a single LCR-Rot with hierarchical attention to handle the output from the Double Rotatory Attention. Table 2 describes the best results of the training accuracy and test accuracy of the new models on two datasets. For the performance of HAABSA-DRA with Type I in SemEval 2015, the training accuracy is rather high however the test accuracy is relatively low, which is a sign of overfitting. On SemEval 2016, the training accuracy and the best test accuracy are close to each other, which can be explained that on SemEval 2016, the data distribution in the train set is similar to the data distribution in the test set. The results of HAABSA-DRA with Type II shows similar pattern as for Type I.

Table 2 Accuracy of Hybrid Approach for Aspect-Based Sentiment Analysis using Double Rotatory Attention model with Type I and Type II

	SemEval 2015	**SemEval 2016**
HAABSA-DRA with Type I		
Train Data	91.39%	94.73%
Test Data	82.41%	88.92%
HAABSA-DRA with Type II		
Train Data	92.88%	94.57%
Test Data	81.70%	89.85%

Undoubtedly, models using imbalanced datasets may produce unreliable results since classes with few samples are more likely to be wrongly marked as other classes during training. To further test the reliability of our model, we calculate the precision of the two models on the given full datasets by the newly designed neural network only. Table 3 displays the obtained precision results. Even for minority classes (e.g., partial datasets with neutral polarity), our model can distinguish them well, which means that an imbalanced training set has a subtle effect on our model.

Table 3 Precision of Hybrid Approach for Aspect-Based Sentiment Analysis using Double Rotatory Attention model with Type I and Type II

	SemEval 2015	**SemEval 2016**
HAABSA-DRA with Type I		
Positive	96.0%	93.0%
Neutral	83.2%	87.5%
Negative	75.3%	72.5%
HAABSA-DRA with Type II		
Positive	84.6%	91.0%
Neutral	66.7%	81.1%
Negative	76.6%	90.6%

Further on, we compare the accuracy of our method with other recent models in the domain of Aspect-Based Sentiment Analysis. Hereby, we define the abbreviations of those models we compare to:

- **Ont+SVM**: A hybrid model that utilizes lexicalized domain ontology sequentially with a Support Vector Machine classifier to determine the polarity of aspect sentiment. (Schouten and Frasincar, 2018).
- **LCR-Rot**: Left-Center-Right Separated Neural Network with Rotatory Attention (Zheng and Xia, 2018).
- **LCR-Rot+hop**: LCR-Rot model with repeating the rotatory attention mechanism for x times. Generally, x is set to 3 based upon (Wallaart and Frasincar, 2019).
- **Ont+LCR-Rot+hop (HAABSA)**: A hybrid method consisting of **Ont** and **LCR-Rot+hop** (Wallaart and Frasincar, 2019).
- **Ont+BERT+LCR-Rot+hop+Hier-Attention (HAABSA++)**: A hybrid method composed of **Ont+LCR-Rot+hop** with Hierarchical Attention, and BERT (Truşcă et al., 2020).
- **BERT+Double LCR-Rot+hop+Type I**: Our method with Type I attention.
- **BERT+Double LCR-Rot+hop+Type II**: Our method with Type II attention.
- **Ont+BERT+Double LCR-Rot+hop+Type I**: A hybrid method consisting of **Ont** and our method with Type I attention.
- **Ont+BERT+Double LCR-Rot+hop+Type II**: A hybrid method consisting of **Ont** and our method with Type II attention.

Ma et al. (2018) develops a hybrid method by extending basic LSTM networks with stacked attention mechanism and recurrent additive network. Reddy et al. (2020) fine-tunes a subset of weights of the model built for comparison with BERT and generic word embeddings. Wallaart and Frasincar (2019) applies the HAABSA model and iterates multiple times over a rotatory attention mechanism. Truşcă et al. (2020) implements HAABSA++ model with deep contextual word embeddings and hierarchical attention. Another hybrid approach is proposed by Meškelė and Frasincar (2020), which integrates a lexicalized domain ontology and a regularized neural attention model. Lately, a semantics perception and refinement network with dual gated multichannel convolution is introduced by Song et al. (2021). For a fair comparison, we compare the accuracy on the SemEval 2015 Task 12 Subtack 2, and the SemEval 2016 Task 5 Subtask 1 datasets. Since most of the scientific work models have not demonstrated precision results on these datasets, we can only use accuracy as a yardstick for comparison. The results are listed in Table 4.

An apparent progression is shown by looking at the accuracy improvement of the first seven model variants. Ontology-based rules, multi-hop technology, hierarchical attention, BERT, along with our new methods, have demonstrated their technological innovation and improvement on LCR-Rot. Using purely the neural attention models we propose, the results on SemEval 2015 and SemEval 2016 compare well against other state-of-the-art models, especially for HAABSA-DRA with Type II. However, they did not beat all the models at SevEval 2015. According to Section 4, SemEval 2015 and SemEval 2016 are about the same in terms of polarity and class distribution, with the biggest difference being the size of the dataset (the SemEval 2016 dataset is larger than the SemEval 2015 dataset). We believe that hybrid methods are an excellent choice

Table 4 Comparison of HAABSA-DRA with Type I and II with other methods using accuracy on Restaurant reviews provided by SemEval-2015 and SemEval-2016

MODEL	SemEval2015	SemEval 2016
Ont+SVM (Schouten and Frasincar, 2018)	63.3%	78.3%
LCR-Rot (Zheng and Xia, 2018)	76.6%	84.6%
LCR-Rot+hop (Wallaart and Frasincar, 2019)	78.4%	86.3%
Ont+LCR-Rot+hop (Wallaart and Frasincar, 2019)	80.6%	87.1%
Ont+BERT+LCR-Rot+hop+Hier-Attention (Truşcă et al., 2020)	81.7%	88.0%
BERT+Double LCR-Rot+hop+Hier+Type I	82.4%	88.9%
BERT+Double LCR-Rot+hop+Type II	81.7%	**89.9**%
Ont+BERT+Double LCR-Rot+hop+Type I	82.8%	87.3%
Ont+BERT+Double LCR-Rot+hop+Type II	82.6%	87.5%
Sentic LSTM (Ma et al., 2018)	78.3%	-
BERT-IL Fine-tuned (Reddy et al., 2020)	-	88.7%
ALDONAr (Meškelė and Frasincar, 2020)	83.8%	87.1%
SPRN(BERT) (Song et al., 2021)	**85.3**%	89.4%

Note: "-" means the according accuracy is not provided.

in the case of small training datasets. If the superiority of the backup model is much higher than that of the ontology, we recommend skipping the ontology step. Considering the fact that SemEval 2015 dataset is smaller in size, its training set may not support sufficient training of our neural network. Hence, the importance of a sufficiently rich amount of data should be emphasized for our approach. Our hybrid approach improves accuracy by nearly 1% on SemEval 2015 compared to our new neural network model. It is worth noting that conversely, the hybrid approach performed worse than our backup model in Several 2016. To explain this phenomenon, we clarify the mechanism of ontology used in the hybrid approach. The ontology can make predictions in 53% of the data with an accuracy of 83% on the SemEval 2015 test dataset and 54% of the data with an accuracy of 87% on the SemEval 2016 test dataset. The ontology is only able to predict the easy cases, where (1) the ontology has coverage (the ontology has the text words as concept representations), and (2) is not conflictual sentiment (both positive and negative for an aspect). In SemEval 2015, the small datasets led to the availability of hybrid methods, whereas in SemEval 2016, neural networks did not require ontology to improve model performance.

To explain why HAABSA-DRA is better than the baseline HAABSA++ method, which is taken as our benchmark model, we analyze the differences between attention weights. Figure 7 shows the visualization of a sentence from the SemEval 2016 test dataset, in which all three models make correct predictions (true sentiment and prediction sentiment are “positive”). The analyzed sentence is “and the waiter suggested a perfect sake!” The intensity of orange indicates the importance of the word represented by the attention score. The darker the color, the higher the weight of attention and the more important a word is in sentiment prediction.

The aspect of the sentence is the word “waiter”. And the opinion expression (e.g., “perfect”) expresses a positive polarity and is in the right context. The context on the left is too short and irrelevant to the aspect word. Only the HAABSA-DRA with two single rotatory structures model captures the most indicative sentiment word “perfect” with the highest attention score, leading to a good sentiment prediction. Although the HAABSA-DRA with three-step rotatory with hierarchical attention structure model does not find "perfect" relevant, it assigns the highest attention score to the word "suggested". It is not as direct as "perfect", but it is equally adequate because "suggested" usually means friendly and polite attitude for "waiter". The HAABSA++ model assigns the highest attention score to the punctuation "!" at the end of the sentence to express emotion. However, it is not reliable to judge sentiment using an exclamation mark, because, in real life, an exclamation mark can express both positive and negative emotion, which will lead to a relatively high error rate. All in all, the HAABSA-DRA models show potential in better capturing the more appropriate sentiment words belonging to a specific aspect to judge, especially when sentences are ambiguous.

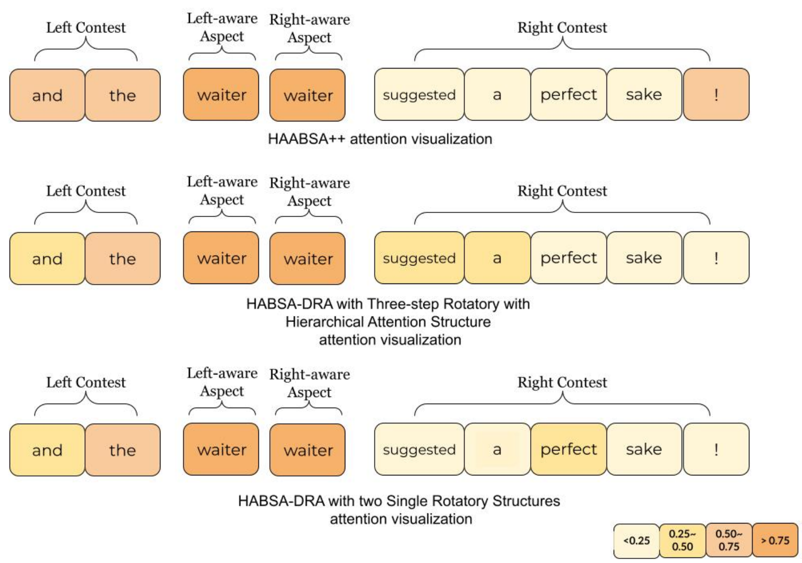

Figure 7: Attention visualizations of the HAABSA++, HAABSA-DRA with Three-step Rotatory with Hierarchical Attention Structure, and HAABSA-DRA with two Single Rotatory Structures models for the phrase 'and the waiter suggested a perfect sake!'

6 Conclusion

In this paper, we employ a hybrid approach for aspect-based sentiment analysis using double rotatory attention with Type I and Type II attention for sentence-level aspect-based sentiment analysis of restaurant reviews. Our proposed ontology-driven hybrid solution further improves the LCR-Rot neural networks backup method by applying a three-stage rotatory attention mechanism with an attention technique. Besides, both HAABSA-DRA models with Type I and II attention re-scale semantic relations via the proposed rotatory attention method and boost up the test accuracy on SemEval 2016. However, the use of the ontology method depends on whether the backup is far superior to the hybrid method. In cases where the backup model is far superior to the hybrid approach, we recommend skipping the ontology steps. Essentially, HAABSA-DRA models with Type I and Type II attention are composed of a double LCR-Rot structure and some additional attention structures, which means different combinations have profound impacts on the generalization ability and results of prediction accuracy. In the future, we would like to evaluate our proposed models on other domains, e.g., laptops and books.

Acknowledgements

We thank Maria Trusca Truşcă and Ximei Yang for helpful comments and contributions during this research.

References

Alani, H., Kim, S., Millard, D. E., Weal, M. J., Hall, W., Lewis, P. H. and Shadbolt, N. R. (2003), 'Automatic ontology-based knowledge extraction from web documents', *IEEE Intelligent Systems* **18**(1), 14–21.

Bengio, Y., Ducharme, R., Vincent, P. and Jauvin, C. (2003), 'A neural probabilistic language model', *Journal of Machine Learning Research* **3**(Feb), 1137–1155.

Bloehdorn, S., Cimiano, P., Hotho, A. and Staab, S. (2005), An ontology-based framework for text mining., *in* 'LDV Forum', Vol. 20, pp. 87–112.

Chen, Z., Mukherjee, A. and Liu, B. (2014), Aspect extraction with automated prior knowledge learning, *in* 'Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)', pp. 347–358.

Coden, A., Gruhl, D., Lewis, N., Mendes, P. N., Nagarajan, M., Ramakrishnan, C. and Welch, S. (2014), Semantic lexicon expansion for concept-based aspect-aware sentiment analysis, *in* 'Proceedings of the Semantic Web Evaluation Challenge', Springer, pp. 34–40.

Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K. and Kuksa, P. (2011), 'Natural language processing (almost) from scratch', *Journal of Machine Learning Research* **12**(ARTICLE), 2493–2537.

Cui, Y., Chen, Z., Wei, S., Wang, S., Liu, T. and Hu, G. (2016), 'Attention-over-attention neural networks for reading comprehension', *arXiv preprint arXiv:1607.04423* .

Das, S. R. and Chen, M. Y. (2007), 'Yahoo! for amazon: Sentiment extraction from small talk on the web', *Management Science* **53**(9), 1375–1388.

Devlin, J., Chang, M.-W., Lee, K. and Toutanova, K. (2019), 'BERT: Pre-training of deep bidirectional transformers for language understanding', *Preceeding of the North American Chapter of the Association for Computational Linguistics: Langage Techology* pp. 4171–4186.

Hu, M. and Liu, B. (2004), Mining and summarizing customer reviews, *in* 'Proceedings of the tenth ACM SIGKDD International Conference on Knowledge discovery and data mining', ACM, pp. 168–177.

Huang, B., Ou, Y. and Carley, K. M. (2018), Aspect level sentiment classification with attention-over-attention neural networks, *in* 'Proceeding of 2018 International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation', Springer, pp. 197–206.

Huang, E. H., Socher, R., Manning, C. D. and Ng, A. Y. (2012), Improving word representations via global context and multiple word prototypes, *in* 'Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)', pp. 873–882.

Kontopoulos, E., Berberidis, C., Dergiades, T. and Bassiliades, N. (2013), 'Ontology-based sentiment analysis of twitter posts', *Expert Systems with Applications* **40**(10), 4065–4074.

LeCun, Y., Bengio, Y. and Hinton, G. (2015), 'Deep learning', *Nature* **521**(7553), 436–444.

Li, H. (2017), 'Deep learning for natural language processing: advantages and challenges', *National Science Review* **5**(1), 24–26.

Li, Z., Raskin, V. and Ramani, K. (2008), 'Developing engineering ontology for information retrieval', *Journal of Computing and Information science in Engineering* **8**, 011003.

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

Liu, Q., Zhang, H., Zeng, Y., Huang, Z. and Wu, Z. (2018), Content attention model for aspect based sentiment analysis, *in* 'Proceedings of the 2018 World Wide Web Conference', ACM, pp. 1023–1032.

Ma, D., Li, S., Zhang, X. and Wang, H. (2017), 'Interactive attention networks for aspect-level sentiment classification', *arXiv preprint arXiv:1709.00893* .

Ma, Y., Peng, H., Khan, T., Cambria, E. and Hussain, A. (2018), 'Sentic lstm: a hybrid network for targeted aspect-based sentiment analysis', *Cognitive Computation* **10**(4), 639–650.

Maas, A., Daly, R. E., Pham, P. T., Huang, D., Ng, A. Y. and Potts, C. (2011), Learning word vectors for sentiment analysis, *in* 'Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies', ACL, pp. 142–150.

Maedche, A. and Staab, S. (2004), Ontology learning, *in* 'Handbook on ontologies', Springer, pp. 173–190.

Meškelė, D. and Frasincar, F. (2019), Aldona: A hybrid solution for sentence-level aspect-based sentiment analysis using a lexicalised domain ontology and a neural attention model, *in* 'Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing', ACM, pp. 2489–2496.

Meškelė, D. and Frasincar, F. (2020), 'Aldonar: A hybrid solution for sentence-level aspect-based sentiment analysis using a lexicalized domain ontology and a regularized neural attention model', *Information Processing & Management* **57**(3), 102211.

Mikolov, T., Chen, K., Corrado, G. and Dean, J. (2013), 'Efficient estimation of word representations in vector space', *Proceedings of Workshop at ICLR* .

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S. and Dean, J. (2013), 'Distributed representations of words and phrases and their compositionality', *27th Annual Conference on Neural Information Processing Systems* **26**, 3111–3119.

Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D. and Riedmiller, M. (2013), 'Playing atari with deep reinforcement learning', *arXiv preprint arXiv:1312.5602* .

Morin, F. and Bengio, Y. (2005), Hierarchical probabilistic neural network language model., *in* 'Aistats', Vol. 5, pp. 246–252.

Pennington, J., Socher, R. and Manning, C. D. (2014), GloVe: Global vectors for word representation, *in* 'Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing', pp. 1532–1543.

Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., Al-Smadi, M., Al-Ayyoub, M., Zhao, Y., Qin, B., De Clercq, O. et al. (2016), Semeval-2016 task 5: Aspect based sentiment analysis, *in* 'Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval 2016)', ACL, pp. 19–30.

Pontiki, M., Galanis, D., Papageorgiou, H., Manandhar, S. and Androutsopoulos, I. (2015), Semeval-2015 task 12: Aspect based sentiment analysis, *in* 'Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)', ACL, pp. 486–495.

Poria, S., Cambria, E. and Gelbukh, A. (2016), 'Aspect extraction for opinion mining with a deep convolutional neural network', *Knowledge-Based Systems* **108**, 42–49.

Reddy, N., Singh, P. and Srivastava, M. M. (2020), 'Does BERT understand sentiment? leveraging comparisons between contextual and non-contextual embeddings to improve aspect-based sentiment models', *arXiv preprint arXiv:2011.11673* .

Ruder, S., Ghaffari, P. and Breslin, J. G. (2016), 'A hierarchical model of reviews for aspect-based sentiment analysis', *arXiv preprint arXiv:1609.02745* .

Schouten, K. and Frasincar, F. (2015), 'Survey on aspect-level sentiment analysis', *IEEE Transactions on Knowledge and Data Engineering* **28**(3), 813–830.

Schouten, K. and Frasincar, F. (2018), Ontology-driven sentiment analysis of product and service aspects, *in* 'Proceedings of the 15th Extended Semantic Web Conference', Springer, pp. 608–623.

Schouten, K., Frasincar, F. and de Jong, F. (2017), Ontology-enhanced aspect-based sentiment analysis, *in* 'Proceedings of the 17th International Conference on Web Engineering', Springer, pp. 302–320.

Schuller, B. and Knaup, T. (2011), Learning and knowledge-based sentiment analysis in movie review key excerpts, *in* 'Toward Autonomous, Adaptive, and Context-Aware Multimodal Interfaces. Theoretical and Practical Issues', Springer, pp. 448–472.

Schwartz, B. and Ward, A. (2004), 'Doing better but feeling worse: The paradox of choice', *Positive psychology in practice* pp. 86–104.

Song, W., Wen, Z., Xiao, Z. and Park, S. C. (2021), 'Semantics perception and refinement network for aspect-based sentiment analysis', *Knowledge-Based Systems* **214**, 106755.

Tang, D., Qin, B., Feng, X. and Liu, T. (2015), 'Effective lstms for target-dependent sentiment classification', *arXiv preprint arXiv:1512.01100* .

Tang, D., Wei, F., Yang, N., Zhou, M., Liu, T. and Qin, B. (2014), Learning sentiment-specific word embedding for twitter sentiment classification, *in* 'Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)', pp. 1555–1565.

Truşcă, M. M., Wassenberg, D., Frasincar, F. and Dekker, R. (2020), A hybrid approach for aspect-based sentiment analysis using deep contextual word embeddings and hierarchical attention, *in* '20th International Conference on Web Engineering', Springer, pp. 365–380.

Vijayaprabakaran, K. and Sathiyamurthy, K. (2020), 'Towards activation function search for long short-term model network: A differential evolution based approach', *Journal of King Saud University-Computer and Information Sciences* .

Wallaart, O. and Frasincar, F. (2019), A hybrid approach for aspect-based sentiment analysis using a lexicalized domain ontology and attentional neural models, *in* 'Proceedings of the 16^{th} Semantic Web Conference', Springer, pp. 363–378.

Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K. et al. (2016), 'Google's neural machine translation system: Bridging the gap between human and machine translation', *arXiv preprint arXiv:1609.08144* .

Yu, J., Zha, Z.-J., Wang, M. and Chua, T.-S. (2011), Aspect ranking: identifying important product aspects from online consumer reviews, *in* 'Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies', ACL, pp. 1496–1505.

Zhang, L., Wang, S. and Liu, B. (2018), 'Deep learning for sentiment analysis: A survey', *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* **8**(4), e1253.

Zhang, L., Zhu, M. and Huang, W. (2009), 'A framework for an ontology-based e-commerce product information retrieval system.', *Journal of ComputerS* **4**(6), 436–443.

Zheng, S. and Xia, R. (2018), 'Left-center-right separated neural network for aspect-based sentiment analysis with rotatory attention', *arXiv preprint arXiv:1802.00892* .

Zhuang, L., Schouten, K. and Frasincar, F. (2020), 'Soba: Semi-automated ontology builder for aspect-based sentiment analysis', *Journal of Web Semantics* **60**, 100544.