

Interactive Learning Approach for Arabic Target-Based Sentiment Analysis

Husamelddin A.M.N Balla , Marisa Llorens Salvador and Sarah Jane Delany

School of Computer Science
Technological University Dublin
Dublin, Ireland

{husamelddin.ballaballa,marisa.llorens,sarahjane.delany}@tudublin.ie

Abstract

Recently, the majority of sentiment analysis researchers focus on target-based sentiment analysis because it delivers in-depth analysis with more accurate results as compared to traditional sentiment analysis. In this paper, we propose an interactive learning approach to tackle a target-based sentiment analysis task for the Arabic language. The proposed IA-LSTM model uses an interactive attention-based mechanism to force the model to focus on different parts (targets) of a sentence. We investigate the ability to use targets, right and left contexts, and model them separately to learn their own representations via interactive modeling. We evaluated our model on two different datasets: Arabic hotel review and Arabic book review datasets. The results demonstrate the effectiveness of using this interactive modeling technique for the Arabic target-based sentiment analysis task. The model obtained accuracy values of 83.10 compared to SOTA models such as AB-LSTM-PC which obtained 82.60 for the same dataset.

1 Introduction

Sentiment analysis (SA) is one of the most prolific research areas in computer sciences, which aims to identify and extract user opinions from reviews. This technique has become an essential part of a wide range of applications in the areas of politics, business, advertising and marketing as it can help in identifying people's opinions towards related targets (Tang et al., 2015). Arabic is considered among the top 4 languages in terms of internet usage (Boudad et al., 2018). With the rapid growth of Arabic web content and low resources for analyzing Arabic opinion mining, the need for accurate Arabic sentiment analysis tools is very necessary.

There are three main classification levels in sentiment analysis: document-level, sentence-level, and target-level. Document-level SA aims to classify the sentiment expressed in the whole document. It

considers the whole document as a basic information unit (talking about one topic). Sentence-level SA aims to classify the sentiment expressed in each sentence. In traditional sentiment analysis, the detailed opinions of all targets of the entity (which are required in many applications) are not provided. To acquire these details, we need to use the target level. Target-based sentiment analysis (TBSA) aims to classify the sentiment with respect to the specific targets of entities. The opinion holders can give different opinions for different targets of the same entity, like in this sentence: "The hotel is clean with good services, but the room was too small". Target-based sentiment analysis is a fine-grained task in sentiment analysis. This kind of fine-grained target-based analysis generally relies on machine learning techniques that require large domain-specific datasets with manual training data (Hu and Liu, 2004).

Target-based sentiment analysis has become more popular in recent research as it delivers more accurate results compared to traditional sentiment analysis. Given a plain text, the trained model is able to detect the targets that were seen in the training set; the context is simply the sequence of words or tokens around the targets. Referring to the previous example, Target1: hotel, Context: clean, good service, Polarity: positive; Target2: room, Context: too small, Polarity: negative. In traditional sentiment analysis (non-target-based SA), this detailed level of analysis is not possible as the analysis is performed at sentence level and hence the entire sentence is classified as either positive or negative.

The research area of Arabic SA is relatively new. Recently, the work on Arabic SA has received a lot of attention and a number of papers on traditional SA have been published in the last couple of years (Shoukry and Rafea, 2012; Duwairi and El-Orfali, 2014; Nabil et al., 2015; Al-Rubaiee et al., 2016). However, research work on Arabic TBSA has not been addressed in depth yet. Most of the existing

TBSA research focuses on English (Xue and Li, 2018; Mowlaei et al., 2020; Ma et al., 2018) with very little work on other languages such as Arabic.

In this paper, we address the task of TBSA in Arabic. The proposed model uses a neural network with an attention mechanism to force the model to attend to the important parts of a sentence. To achieve that, an interactive attention-based long short-term memory network (IA-LSTM) with an interaction technique is used to capture important information related to a given target.

Previous work in Arabic target-based SA used deep learning represented in Recurrent Neural Network (RNN) and LSTM models and developed several methods aiming to model contexts through the generation of target-based representations (Ruder et al., 2016; Tamchyna and Veselovská, 2016; Al-Smadi et al., 2019). However, the modeling of interactive targets with contexts using attention mechanisms was not addressed.

The model proposed in this paper (IA-LSTM) is based on a model (IAN) proposed for English language TBSA in a previously published paper (Ma et al., 2017) that addresses the separately modeled targets and contexts that jointly interact with each other with an attention mechanism. However, the left context was ignored in the previously proposed approach in (Ma et al., 2017). Given the fact that Arabic language is written from right to left, the addition of the left context may improve the performance as the most of the opinionated words come after the target (left context). The proposed model (IA-LSTM) is obtained by adding the left context to the IAN model from (Ma et al., 2017), to end up with a model that considers the three main elements (left context, target, and right context). Code for the proposed model is publicly available on Github ¹.

The rest of this paper is organized as follows: section 2 presents the related work of SA and TBSA in English and Arabic. Section 3 discusses the methodology used for the proposed model in this paper. Section 4 describes the datasets used for evaluations. Section 5 explains the baselines, evaluations, the experimental results and analysis. Section 6 concludes this paper.

2 Related Work

The research work on Arabic TBSA has few examples as the published papers are limited, so, the

first subsection of related work mainly focus on traditional Arabic SA while the second subsection covers the published work on TBSA on both Arabic and other languages.

2.1 Arabic SA

The early attempts on Arabic SA relied on the methods applied for English SA as it is more mature with rich resources in terms of SA. An example of this, Ahmad (2006) employed a rule-based approach that was originally designed for English. Almas and Ahmad (2007) modified the approach to accept other languages such as English and Urdu. Those two attempts used financial news datasets. The results from these experiments were similar across all of languages tested.

Continuing in a business-oriented and rule-based approach, Elhawary and Elfeky (2010) performed experiments on Arabic SA. The authors addressed the problem of SA using large data and MapReduce in an attempt to enhance the performance. Another work following a similar approach was proposed by Farra et al. (2010), which used a pre-compiled lexicon to improve the performance on both sentence-level and document-level SA. However, there were no significant improvements in general performance, except at sentence level SA with slight improvement.

Work on Arabic SA or any other language always requires datasets to test the used approach. As Arabic is a language with low resources, very few datasets have been created for Arabic SA over the past years. AWATIF Abdul-Mageed and Diab (2012) is one example of the existing datasets.

2.2 Target Based Sentiment Analysis (TBSA)

One of the earliest papers in TBSA was for English, and used a frequency-based approach proposed by Hu and Liu (2004). The basic idea behind this approach is counting the nouns in the text and considering the most frequently mentioned ones as targets. The authors also tried to avoid the error of incorrectly identifying infrequent nouns as targets by using the nearest opinion words. This approach was also used in several other papers (Qiu et al., 2009; Zhuang et al., 2006). To enhance this approach, Popescu and Etzioni (2007) proposed using a technique named “part of relationship” to eliminate the frequent nouns that are incorrectly identified as targets.

TBSA is more challenging than the traditional task of sentiment analysis (non-target-based) be-

¹<https://github.com/HUSTUD/IA-LSTM>

cause the model needs to include the impact of the context words on the target. A deep learning approach in this task can be performed by representing context, generating a target representation, and then detecting the important parts of the sentence (i.e. the targets). Again, RNNs have proven their competitive performance in this task in terms of capturing long-term dependency in sentences and general semantic classification. Moreover, the best RNN performers are the ones that include attention or memory networks. This shows that the models can learn how to concentrate on different parts of the sentence with an attention weight aggregated from a lower level to classify targets and opinion words and the link between them. Several researchers adopted an attention mechanism for this task in English (Gers et al., 2000; Song et al., 2019; Li et al., 2018). In general, there are not many techniques using deep learning for Arabic sentiment analysis and, in particular, TBSA (Dahou et al., 2016; Ruder et al., 2016; Tamchyna and Veselovská, 2016; Al-Smadi et al., 2019).

The International Workshop on Semantic Evaluation (SemEval), one of the most significant events in natural language processing (NLP) research, is concerned with the evaluation of computational semantic analysis systems. A special event focused on TBSA was organized in 2014, 2015 and 2016. In this workshop, Al-Smadi et al. (2018) proposed Arabic TBSA paper which performs a comparison between a deep neural network and SVM models. The authors used an RNN framework named Deeplearning4j that provides a set of implementations for different deep neural network algorithms. They evaluated their models on the Arabic hotel review dataset (Mohammad et al., 2016). Their deep neural network model outperformed the SVM model in accuracy of sentiment polarity while SVM outperformed the deep learning model in target extraction. They also proposed another study (Mohammad et al., 2016), which was part of the SemEval-TBSA 2016 competition. In this study they created a hotel review dataset with baselines obtained by using the SVM model with only unigram feature.

Focusing on Arabic TBSA and using the same dataset that this research is using (Arabic hotel reviews), three papers were recently published (Ruder et al., 2016; Tamchyna and Veselovská, 2016; Al-Smadi et al., 2019). Ruder et al. (2016) proposed a deep learning-based approach

(INSIGHT-1) for multi-lingual TBSA as one of SemEval-2016 participants, which used a convolutional neural network for target extraction and sentiment analysis. Using the Arabic hotel review dataset their model outperformed the other participants in the workshop. Another approach proposed by Tamchyna and Veselovská (2016) for the same task used an RNN-based binary classifier for the task of target-category identification. Their model was trained using word embedding features and achieved good performance as the second rank after (Ruder et al., 2016).

A recent attempt by Al-Smadi et al. (2019) used targets and context embeddings in their proposed model (AB-LSTM-PC). The approach models the context words via LSTM networks and then combines the word's hidden states with target embeddings to generate the attention vectors. In addition, to further strengthen the effect of target embeddings, the model appends target embeddings, with each word embedding vector forming the context. This is used to produce the final representation for TBSA. Their model performance was the highest among all previously published papers evaluated on the same dataset we are using in this research.

An approach (IAN) for English TBSA using interactive targets and context representations was proposed in Ma et al. (2017). The model uses attention mechanisms to concatenate the separately modeled targets and context as final representation before it is fed to the softmax layer. This model only considers the right context of targets, ignoring the left context.

All the research mentioned in this section ignores the the left and right context of targets and the interactions between them. Using these extra features can increase the amount of information about the context by providing a more comprehensive approach to context. In this paper, we propose an interactive learning approach based on (Ma et al., 2017) to tackle sentiment polarity identification for the Arabic language which includes the use of left and right context.

3 Proposed Approach

Recurrent neural networks (RNNs) are deep learning neural networks designed specifically to learn sequences of data and are mainly used for textual data classification. The learning process is done at hidden recurrent nodes depending on their previous layers of nodes. However, RNNs suffer from the

vanishing gradient problem when handling long sequences of data. Bi-directional long short-term memory (Bi-LSTM) (Schuster and Paliwal, 1997) was proposed as a solution for this problem and have proven to be efficient in many NLP-related problems. In contrast to the standard RNNs, Bi-LSTM units have a major role in extracting and learning important features out of the input or computed data and keeping the computed values as long as they are needed in the memory vector.

In this research, we focus on target-based sentiment polarity classification considering the right and left contexts. The enhancement of sentiment classification performance can be achieved by considering the targets and their contexts. The good performance relies on simultaneously modeling targets and contexts precisely. Targets and contexts can influence the representation of each other. For example, the target word “hotel” can be naturally associated with the context word “clean” and vice versa. Therefore, targets and contexts can be modeled individually but learn from their interactions. In the input text, each word usually has its own contribution or importance, which is different from other words in the final representation for sentiment analysis. For instance, the importance of the word “room” is higher in the representation of the target “room price”, which is described by “expensive”. Therefore, in the proposed model, the attention weights for both targets and contexts are computed to respectively capture their important information.

The proposed model is an interactive attention LSTM-based (IA-LSTM) model which employs long short-term memory networks (LSTM) and attention mechanisms. To get important information from the left and right context, the model uses an attention mechanism associated with a target then computes context representation for sentiment polarity identification. In addition, the proposed model makes use of the interactive information from the word’s context to supervise the target modeling. Finally, the model concatenates the attended left context representations, target representations, and right context representations, then uses them to predict the sentiment polarity (see Fig. 1).

Following the model notation in Fig. 1, assume that a left context consists of k words $[w_{LC1}, w_{LC2}, \dots, w_{LCk}]$, a target consists of m words $[w_{t1}, w_{t2}, \dots, w_{tm}]$, and a right context consists of n words $[w_{RC1}, w_{RC2}, \dots, w_{RCn}]$. We use pre-trained word embeddings for word representation of contexts

and targets. Then, since there is a strong dependence between words in a sentence, the LSTM network is used to learn the hidden word semantics as LSTM is good at learning long-term dependencies. Next, LSTM produces the hidden states $[h_{LC1}, h_{LC2}, \dots, h_{LCk}]$ for the left context words, $[h_{t1}, h_{t2}, \dots, h_{tm}]$ for the target words, and $[h_{RC1}, h_{RC2}, \dots, h_{RCn}]$ for the right context as word representations. Then, the model calculates the average of the hidden states to get the initial representations of the contexts and target.

The attention mechanism in the proposed model is adopted by using the initial representations of the contexts and target as input to help in selecting important information for classifying the sentiment polarity. To form the attention mechanism, we use the average as well as the last hidden state output of target and contexts to capture an abstract representations of the input sequence. The use of the average value is to form the initial representation of the other component (target or context) and combine it with the last hidden state of the current component. We found this is the best way to form the attention mechanism and it help to reduce the noise and the sparse information associated with the Arabic language (Elnagar et al., 2020).

The attention process is described by considering the left context, target, and the right context, as shown in Fig. 1. Using the left context words representation $[h_{LC1}, h_{LC2}, \dots, h_{LCk}]$ and the average of hidden states of target representation T_{avg} , the model computes the left attention vector $L\alpha_i$. Similarly, the model computes the target attention vector β_i by using the average of left context words LC_{avg} , the target words representation $[h_{t1}, h_{t2}, \dots, h_{tm}]$, and the average of the right context vector RC_{avg} . The same technique is followed for the right attention vector $R\alpha_i$ obtained using the right context words representation $[h_{RC1}, h_{RC2}, \dots, h_{RCn}]$ and the average of hidden states of target representation T_{avg} . Then, the attention vectors $L\alpha_i, \beta_i$, and $R\alpha_i$ are used to obtain the word attention weights and concatenated as one vector before being fed to the softmax classifier.

Formally, given the word embeddings of the contexts and target, the hidden state representations of the contexts (left and right) and target are $[h_{LC1}, h_{LC2}, \dots, h_{LCk}]$, $[h_{RC1}, h_{RC2}, \dots, h_{RCn}]$ and $[h_{t1}, h_{t2}, \dots, h_{tm}]$ respectively. The initial representation of contexts (left - equation (1) and right - equation (2)) and targets equation (3) is obtained by calculating

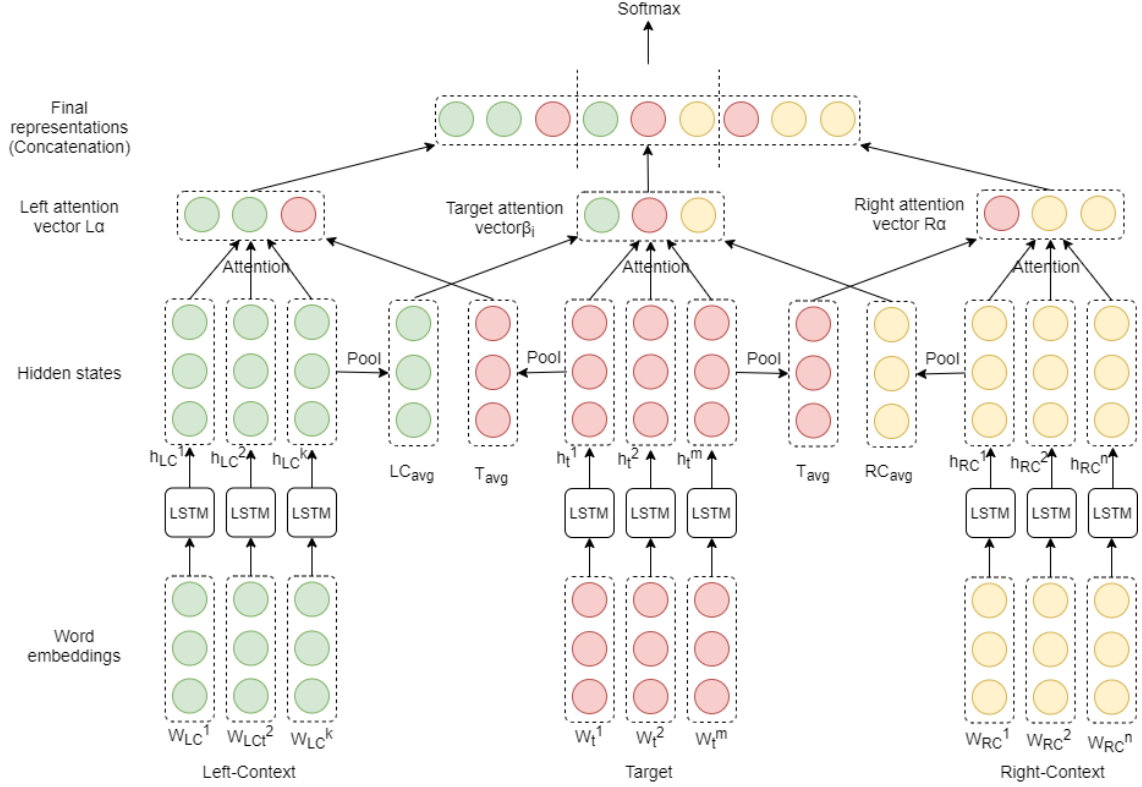


Figure 1: The general architecture of the proposed model for Arabic target-based sentiment analysis.

the average of each as follows:

$$LC_{avg} = \sum_{i=1}^k \frac{h_{LC}^i}{k} \quad (1)$$

$$RC_{avg} = \sum_{i=1}^n \frac{h_{RC}^i}{n} \quad (2)$$

$$T_{avg} = \sum_{i=1}^m \frac{h_T^i}{m} \quad (3)$$

To make the model focus on the important parts of the representations, an attention mechanism that employs the initial representations of the contexts and target is used. The target influence on the left context and the influence of the left context on the target is considered, as well as the influence of the target on the right context and the influence of the right context on the target. The left attention vector $L\alpha_i$ is generated using both the left context and target representations as in equation (4):

$$L\alpha_i = \frac{e^{\gamma(h_{LC}^i, T_{avg})}}{\sum_{j=1}^k e^{\gamma(h_{LC}^j, T_{avg})}} \quad (4)$$

Where γ is a score function that computes the importance of h_{LC}^i and h_{RC}^i in the left and right

context respectively. This function is defined as equation (5):

$$\gamma(h_{LC}^j, T_{avg}) = \tanh(h_{LC}^j \cdot w \cdot T_{avg}^t + b) \quad (5)$$

Where w is a weight matrix, b is a bias, and T_{avg}^t is a transpose of T_{avg} . The right attention vector $R\alpha_i$ is generated using both the right context and target representations as equation (6):

$$R\alpha_i = \frac{e^{\gamma(h_{RC}^i, T_{avg})}}{\sum_{j=1}^n e^{\gamma(h_{RC}^j, T_{avg})}} \quad (6)$$

Similarly, the target attention vector β_i is calculated using all of the right context, left context, and the target as (7):

$$\beta_i = \frac{e^{\gamma(h_T^i, LC_{avg})}}{\sum_{j=1}^m e^{\gamma(h_T^j, LC_{avg})}} + \frac{e^{\gamma(h_T^i, RC_{avg})}}{\sum_{j=1}^m e^{\gamma(h_T^j, RC_{avg})}} \quad (7)$$

The representations of the left context, target, and right context are computed as in equations (8), (9), (10) respectively:

$$C_1 = \sum_{i=1}^k L\alpha_i h_{LC}^i \quad (8)$$

$$T = \sum_{i=1}^m \beta_i h_{RC}^i h_{LC}^i \quad (9)$$

$$C_r = \sum_{i=1}^n R\alpha_i h_{RC}^i \quad (10)$$

The final representation is done by concatenating the representations of the three components of the left context, target, and the right context (C_l , T , and C_r respectively) as one vector (F). Finally, the prediction of sentiment polarity (positive, neutral, negative) is done by using the softmax layer as in equation (11):

$$x = \tanh(w_1 \cdot F + b_1) \quad (11)$$

Where w_1 is the weight matrix and b_1 is the bias.

3.1 Hyperparameters and Model Training

The pre-trained Arabic word embeddings ‘‘AraVec’’ (Soliman et al., 2017) was used for the target and the context word embeddings with a dimension of 300 nodes. AraVec is an open-source project which provides free-to-use Arabic word embeddings trained on more than 3 billion words from web pages and Wikipedia. A uniform β distribution $U(-0.1; 0.1)$ was used to initialize all out-of-vocabulary words and weights. The model employs the momentum optimization algorithm (Qian, 1999) to train the parameters, which adds a fraction of the update vector in the preceding step to the current update vector. The dropout rate is set to 0.5, and the normalization coefficient $L2$ in the objective function is set to 10^{-5} .

4 Datasets

There are two benchmark datasets were used to evaluate the proposed approach, Arabic hotel reviews and Arabic book reviews datasets.

4.1 Arabic Hotel Reviews

This dataset was part of the work proposed in task 5 of the Sem-Eval 2016 workshop on target-based sentiment analysis (Pontiki et al., 2016; Mohamad et al., 2016). The dataset contains the 24,028 TBSA annotated tuples provided (19,226 tuples for training and 4,802 tuples for testing). For the sake of generalization and to avoid the single dialect problem, the original dataset was collected from well-known different Hotels’ booking websites such as Booking.com, TripAdvisor.com. The

selected reviews in the datasets belongs to Hotels from different Arabian cities in different countries such as Dubai, Beirut, Amman, Mecca, etc. In addition, the dataset was annotated on both text-level (2,291 reviews’ texts) and sentence-level (6,029 annotated sentences). In this research, we consider only the sentence-level tasks. This is a manually annotated dataset, whereas, for each sentence, a tuple of target category, opinion target expression, and target polarity were annotated. The sentiment polarity labels (positive, negative, neutral) were used to annotate the polarity of each target or category.

4.2 Arabic Book Reviews

This dataset was provided by (Al-Smadi et al., 2015) as a benchmark Human Annotated Arabic Dataset (HAAD). HAAD is a book review dataset in Arabic, which has been constructed and annotated by humans, taking into account the target terms and their polarities. For each review sentence, a tuple consisting of an target-category, target-category polarity, target-term, and target-term polarity was extracted and annotated. A sentiment polarity (positive, negative, conflict, neutral) was used to represent both the target-category and target-term sentiment polarity based on the annotated sentences. This dataset consists of 1513 Arabic book reviews annotated with aspect terms, aspect term polarity, aspect category, and aspect category polarity. In our experiments, we considered the labels positive, negative, and neutral excluding conflict label.

Lastly, both datasets were proposed as three separated machine-readable XML format files annotated training, test, and gold test with polarity distribution in Table 1. In addition, one sentence can contains more than one target which can be assigned with different polarity label. SemEval-TBSA designed a specific way of evaluations for all of the models evaluated on the associated datasets. F1 score is usually used as a metric for category and target execution. However, the metric used for sentiment polarity in the literature review is the accuracy of each model. A sample of Arabic hotel reviews dataset is shown in Fig. 2.

5 Evaluation

We evaluated the proposed model on the Arabic hotel review and the Arabic book review. The evaluation is done by computing the accuracy of the polarity. The used accuracy metric defined as the

```

<sentence id="6:0">
  <text>فندق حلو جدا جدا جميل وعائلي ومناسب</text> Nice and convenient hotel
  <Opinions>
    <Opinion target="فندق" polarity="positive" from="0" to="4"/>
  </Opinions>
  hotel
</sentence>

```

Figure 2: Sample of Arabic hotel reviews dataset.

Dataset	File	Positive	Negative	Neutral
Hotel R	Train	5819	3141	660
Hotel R	Test	1426	784	162
Book R	Train	1252	855	126
Book R	Test	124	432	22

Table 1: Polarity distribution in both Hotel and Book reviews datasets

number of correctly predicted polarity labels of the (gold) targets, divided by the total number of the gold targets. The test set contains sentences without polarity labels which is expected to be predicted by the model and compared with the labels in the test gold set containing the same sentences. We compared the proposed model performance with the related work and SOTA approaches stated as LSTM (Gers et al., 2000), IAN (Ma et al., 2017), INSIGHT-1 (Ruder et al., 2016), and AB-LSTM-PC (Al-Smadi et al., 2019) evaluated on the same Arabic hotel review dataset. For the Arabic book review dataset, the proposed model is compared with a subset of these approaches: LSTM (Gers et al., 2000) and IAN (Ma et al., 2017). In addition, the baselines labelled as “Baseline” in Table 2 are the baselines provided with the published work of dataset creation. The baseline for the Arabic hotel review dataset was obtained by using SVM with N-grams as features. The baseline for the Arabic book reviews was obtained by using a frequency approach.

5.1 Compared Models

The models used in the comparison in this paper are as follows: (i) LSTM uses one LSTM network to model the context and get the hidden state of each word. The average value of all the hidden states is obtained as the final representation and fed to a softmax function to estimate the probability of each sentiment label (Gers et al., 2000). (ii) IAN uses an interactive attention-based LSTM considering the target and right context only (ignoring the left context) (Ma et al., 2017). (iii) INSIGHT-1 uses a

convolutional neural network (CNN). The model concatenates the target vector with each word embedding and then applies a convolution over it to identify the sentiment polarity (Ruder et al., 2016). (iv) AB-LSTM-PC uses an attention-based LSTM by adding the target embedding in the input. To generate the attention vectors, the approach models the context using LSTM networks and combines the hidden states with the target embeddings (Al-Smadi et al., 2019).

For INSIGHT-1 and AB-LSTM-PC, we report the previously published results without implementing the models as they were evaluated on the same dataset. The rest of the compared models (LSTM and IAN) were implemented and evaluated on the datasets.

5.2 Experimental Results and Analysis

Table 2 shows the proposed model’s performance compared with other models. The worst performance in this table was for the LSTM model. This is most likely due to the fact that it does not make use of the attention mechanism, confirming findings in previous research (Jiang et al., 2011). Both the AB-LSTM-PC model (SOTA) and INSIGHT-1 have similar performance and outperform the LSTM model, confirming that the attention mechanism enhances the ability to identify sentiment polarity. Unlike IAN in (Ma et al., 2017), our proposed IA-LSTM model (represented by “IA-LSTM” in the table) takes a further step towards confirming the importance of considering the targets and contexts in the learning process interactively. As shown in Table 2, the IA-LSTM model achieved the highest performance, outperforming all baselines and the other approaches. This enhancement can be explained by the fact that our model uses three connected attention networks to model the target and contexts. Using such a design, the model can effectively learn the representations of targets and contexts, which can jointly enhance the overall performance of target-based sentiment

Model	Hotel R	Book R
LSTM	74.26	71.50
INSIGHT-1	82.7	-
AB-LSTM-PC	82.60	-
IAN	81.90	78.96
IA-LSTM	83.10	80.82
Baseline	76.40	29.70

Table 2: Accuracy comparison between the proposed IA-LSTM model performance for the TBSA task and the baseline results based on SVM along with n-gram features and approaches from related work.

Dataset	Positive	Negative	Neutral
Hotel R	89.86	83.29	76.14
Book R	86.42	87.93	65.82

Table 3: Accuracy for each class in both Hotel reviews and Book reviews datasets

classification. As we can notice from Table 2, different approaches show that the more attention that is paid to targets, the higher the accuracy.

As we can notice for Table 1, the class distribution is imbalanced in both Hotel review and Book review datasets. Therefore, we calculate the accuracy for each polarity class in Table 3. A lower accuracy obtained by class "Neutral" as it has the lowest amount of data and a higher accuracy obtained by class "Positive".

Lastly, we verified the effectiveness of using targets and interaction with the left and right contexts in modeling the attention mechanism and the results are displayed in Table 4. The first model (labelled *No-interaction*) completely ignored the interaction between the targets and contexts. This model uses three LSTM networks to learn the representations of the left context, target, and right context in their own local attentions without any interaction. In the second model (labelled *Right-side interaction*), we showed the impact of using the right-side of the interaction mechanism that enables the target to interact with the right contexts. This model uses two LSTM networks to learn the representations of the target and right context and interact with the right context only. Similarly, in the third model (labelled *Left-side interaction*) the target interacts with the left context only. The fourth model (labelled *Full-model*) uses the target interaction with both left and right contexts. As shown in Table 4, a lower performance was achieved by the *No-interaction* model. Using

Model	Hotel R	Book R
No-interaction	76.75	74.64
Right-side interaction	78.21	75.32
Left-side interaction	81.90	78.96
Full-model	83.10	80.82

Table 4: Accuracy Analysis of the proposed IA-LSTM model in terms of using the target and interaction techniques.

two-side interaction showed some improvement in the performance specially the left-side as Arabic is written from right to left which means most of the opinion words comes on the left-side after the target. The best performance was achieved by the *Full-model*, as we expected, where using the target interactions with the left and right contexts is fully considered, which enhances the overall model performance. Therefore, from this table, we can observe that the interaction of the target between the left and right contexts can contribute greatly in enhancing TBSA. The performance investigation of using target-attention only was well represented by the AB-LSTM-PC model in Table 2 which we avoid to repeat it in our comparison in Table 4.

6 Conclusion

In this paper, we proposed a deep learning-based approach to tackle Arabic TBSA. The proposed model "IA-LSTM" uses an interactive attention-based technique for the task. The main idea of the proposed IA-LSTM model is to use three attention networks to interactively model the target and contexts (left and right). It is based on previous work in English that uses only the right context but given the way Arabic is written from right to left, the addition of the left context provides important context information. The model can focus on the important parts in the sentence and identify the sentiment polarity. The proposed approach was evaluated on two different datasets: Arabic hotel reviews and book reviews. Experiments verify that the proposed approach outperforms the baselines and other approaches evaluated on the same datasets. Implementing the interactive attention-based model demonstrated that the model can learn effective features for targets and contexts (left and right) and enhance the performance of Arabic TBSA.

References

- Muhammad Abdul-Mageed and Mona T Diab. 2012. Awatif: A multi-genre corpus for modern standard arabic subjectivity and sentiment analysis. In *LREC*, volume 515, pages 3907–3914.
- Khurshid Ahmad. 2006. Multi-lingual sentiment analysis of financial news streams. *PoS*, page 001.
- Hamed Al-Rubaiee, Renxi Qiu, and Dayou Li. 2016. Identifying mubasher software products through sentiment analysis of arabic tweets. In *2016 International Conference on Industrial Informatics and Computer Systems (CIICS)*, pages 1–6. IEEE.
- Mohammad Al-Smadi, Omar Qawasmeh, Mahmoud Al-Ayyoub, Yaser Jararweh, and Brij Gupta. 2018. Deep recurrent neural network vs. support vector machine for aspect-based sentiment analysis of arabic hotels’ reviews. *Journal of computational science*, 27:386–393.
- Mohammad Al-Smadi, Omar Qawasmeh, Bashar Talafha, and Muhannad Quwaider. 2015. Human annotated arabic dataset of book reviews for aspect based sentiment analysis. In *2015 3rd International Conference on Future Internet of Things and Cloud*, pages 726–730. IEEE.
- Mohammad Al-Smadi, Bashar Talafha, Mahmoud Al-Ayyoub, and Yaser Jararweh. 2019. Using long short-term memory deep neural networks for aspect-based sentiment analysis of arabic reviews. *International Journal of Machine Learning and Cybernetics*, 10(8):2163–2175.
- Yousif Almas and Khurshid Ahmad. 2007. A note on extracting ‘sentiments’ in financial news in english, arabic & urdu. In *The Second Workshop on Computational Approaches to Arabic Script-based Languages*, pages 1–12.
- Naaima Boudad, Rdouan Faizi, Rachid Oulad Haj Thami, and Raddouane Chiheb. 2018. Sentiment analysis in arabic: A review of the literature. *Ain Shams Engineering Journal*, 9(4):2479–2490.
- Abdelghani Dahou, Shengwu Xiong, Junwei Zhou, Mohamed Houcine Haddoud, and Pengfei Duan. 2016. Word embeddings and convolutional neural network for arabic sentiment classification. In *Proceedings of coling 2016, the 26th international conference on computational linguistics: Technical papers*, pages 2418–2427.
- Rehab Duwairi and Mahmoud El-Orfali. 2014. A study of the effects of preprocessing strategies on sentiment analysis for arabic text. *Journal of Information Science*, 40(4):501–513.
- Mohamed Elhawary and Mohamed Elfeky. 2010. Mining arabic business reviews. In *2010 IEEE international conference on data mining workshops*, pages 1108–1113. IEEE.
- Ashraf Elnagar, Ridhwan Al-Debsi, and Omar Einea. 2020. Arabic text classification using deep learning models. *Information Processing & Management*, 57(1):102121.
- Noura Farra, Elie Challita, Rawad Abou Assi, and Hazem Hajj. 2010. Sentence-level and document-level sentiment mining for arabic texts. In *2010 IEEE international conference on data mining workshops*, pages 1114–1119. IEEE.
- Felix A Gers, Jürgen A Schmidhuber, and Fred A Cummins. 2000. Learning to forget: Continual prediction with lstm. *Neural Computation*, 12(10):2451–2471.
- Minqing Hu and Bing Liu. 2004. Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 168–177.
- Long Jiang, Mo Yu, Ming Zhou, Xiaohua Liu, and Tiejun Zhao. 2011. Target-dependent twitter sentiment classification. In *Proceedings of the 49th annual meeting of the association for computational linguistics: human language technologies*, pages 151–160.
- Xin Li, Lidong Bing, Piji Li, Wai Lam, and Zhimou Yang. 2018. Aspect term extraction with history attention and selective transformation. *arXiv preprint arXiv:1805.00760*.
- Dehong Ma, Sujian Li, Xiaodong Zhang, and Houfeng Wang. 2017. Interactive attention networks for aspect-level sentiment classification. *arXiv preprint arXiv:1709.00893*.
- Yukun Ma, Haiyun Peng, and Erik Cambria. 2018. Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive lstm. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32.
- AL-Smadi Mohammad, Omar Qwasmeh, Bashar Talafha, Mahmoud Al-Ayyoub, Yaser Jararweh, and Elhadj Benkhelifa. 2016. An enhanced framework for aspect-based sentiment analysis of hotels’ reviews: Arabic reviews case study. In *2016 11th International Conference for Internet Technology and Secured Transactions (ICITST)*, pages 98–103. IEEE.
- Mohammad Erfan Mowlaei, Mohammad Saniee Abadeh, and Hamidreza Keshavarz. 2020. Aspect-based sentiment analysis using adaptive aspect-based lexicons. *Expert Systems with Applications*, 148:113234.
- Mahmoud Nabil, Mohamed Aly, and Amir Atiya. 2015. Astd: Arabic sentiment tweets dataset. In *Proceedings of the 2015 conference on empirical methods in natural language processing*, pages 2515–2519.

- Maria Pontiki, Dimitrios Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, Mohammad Al-Smadi, Mahmoud Al-Ayyoub, Yanyan Zhao, Bing Qin, Orphée De Clercq, et al. 2016. Semeval-2016 task 5: Aspect based sentiment analysis. In *International workshop on semantic evaluation*, pages 19–30.
- Ana-Maria Popescu and Oren Etzioni. 2007. Extracting product features and opinions from reviews. In *Natural language processing and text mining*, pages 9–28. Springer.
- Ning Qian. 1999. On the momentum term in gradient descent learning algorithms. *Neural networks*, 12(1):145–151.
- Guang Qiu, Bing Liu, Jiajun Bu, and Chun Chen. 2009. Expanding domain sentiment lexicon through double propagation. In *IJCAI*, volume 9, pages 1199–1204. Citeseer.
- Sebastian Ruder, Parsa Ghaffari, and John G Breslin. 2016. Insight-1 at semeval-2016 task 5: Deep learning for multilingual aspect-based sentiment analysis. *arXiv preprint arXiv:1609.02748*.
- Mike Schuster and Kuldip K Paliwal. 1997. Bidirectional recurrent neural networks. *IEEE transactions on Signal Processing*, 45(11):2673–2681.
- Amira Shoukry and Ahmed Rafea. 2012. Sentence-level arabic sentiment analysis. In *2012 International Conference on Collaboration Technologies and Systems (CTS)*, pages 546–550. IEEE.
- Abu Bakr Soliman, Kareem Eissa, and Samhaa R El-Beltagy. 2017. Aravec: A set of arabic word embedding models for use in arabic nlp. *Procedia Computer Science*, 117:256–265.
- Youwei Song, Jiahai Wang, Tao Jiang, Zhiyue Liu, and Yanghui Rao. 2019. Attentional encoder network for targeted sentiment classification. *arXiv preprint arXiv:1902.09314*.
- Aleš Tamchyna and Kateřina Veselovská. 2016. Ufal at semeval-2016 task 5: recurrent neural networks for sentence classification. In *Proceedings of the 10th international workshop on semantic evaluation (SEMEVAL-2016)*, pages 367–371.
- Duyu Tang, Bing Qin, Xiaocheng Feng, and Ting Liu. 2015. Effective lstms for target-dependent sentiment classification. *arXiv preprint arXiv:1512.01100*.
- Wei Xue and Tao Li. 2018. Aspect based sentiment analysis with gated convolutional networks. *arXiv preprint arXiv:1805.07043*.
- Li Zhuang, Feng Jing, and Xiao-Yan Zhu. 2006. Movie review mining and summarization. In *Proceedings of the 15th ACM international conference on Information and knowledge management*, pages 43–50.