

Article

BERT-Based Joint Model for Aspect Term Extraction and Aspect Polarity Detection in Arabic Text

Hasna Chouikhi ¹, Mohammed Alsuhaibani ^{2,*} and Fethi Jarray ^{1,3}¹ LIMTIC Laboratory, UTM University, Tunis 1068, Tunisia² Department of Computer Science, College of Computer, Qassim University, Buraydah 52571, Saudi Arabia³ Higher Institute of Computer Science of Medenine, Gabes University, Medenine 6029, Tunisia

* Correspondence: m.suhibani@qu.edu.sa

Abstract: Aspect-based sentiment analysis (ABSA) is a method used to identify the aspects discussed in a given text and determine the sentiment expressed towards each aspect. This can help provide a more fine-grained understanding of the opinions expressed in the text. The majority of Arabic ABSA techniques in use today significantly rely on repeated pre-processing and feature-engineering operations, as well as the use of outside resources (e.g., lexicons). In essence, there is a significant research gap in NLP with regard to the use of transfer learning (TL) techniques and language models for aspect term extraction (ATE) and aspect polarity detection (APD) in Arabic text. While TL has proven to be an effective approach for a variety of NLP tasks in other languages, its use in the context of Arabic has been relatively under-explored. This paper aims to address this gap by presenting a TL-based approach for ATE and APD in Arabic, leveraging the knowledge and capabilities of previously trained language models. The Arabic base (Arabic version) of the BERT model serves as the foundation for the suggested models. Different BERT implementations are also contrasted. A reference ABSA dataset was used for the experiments (HAAD dataset). The experimental results demonstrate that our models surpass the baseline model and previously proposed approaches.

Keywords: Arabic aspect term extraction (ATE); aspect sentiment polarity (ASP); Arabic aspect-based sentiment analysis (AABSA); Arabic BERT; CRF; natural language processing systems; neural networks; deep learning; text processing



Citation: Chouikhi, H.; Alsuhaibani, M.; Jarray, F. BERT-Based Joint Model for Aspect Term Extraction and Aspect Polarity Detection in Arabic Text. *Electronics* **2023**, *12*, 515.
<https://doi.org/10.3390/electronics12030515>

Academic Editors: Habib Hamam, Ateeq Ur Rehman and Mohamed Tahar Ben Othman

Received: 12 December 2022

Revised: 11 January 2023

Accepted: 12 January 2023

Published: 19 January 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Sentiment analysis (SA) is a popular research topic in natural language processing (NLP). It has recently gained a lot of attention due to massive online polls on services and products and the growth of social media data. The three levels of sentiment analysis are document, sentence, and aspect levels. The opinion at the document and sentence levels is about the entirety of the given content (document or sentence), and it is frequently insufficient to express an opinion on the particular issues raised in the text. Aspect-based sentiment analysis, or ABSA for short, is a detailed approach to predicting sentiment that considers specific entities, such as goods, services, events, or organizations, in a particular domain. This type of analysis goes beyond the capabilities of traditional sentiment analysis by providing a more nuanced understanding of the sentiment expressed in a text [1]. High-level NLP conferences and workshops, including SemEval, have made ABSA one of their primary topics due to its importance and impact on the field.

In fact, the ABSA task was divided into four subtasks for SemEval 2014 task 4. First, there is aspect term extraction (ATE), where the aim is to take out features or aspects of services, goods, or topics that have been discussed in a phrase or a sentence. For example, “This phone’s camera is quite powerful”. The reviewer assessed the phone’s camera in this clause. Similar to named entity recognition (NER), aspect term extraction (ATE) tasks are viewed as sequence-marking problems. Hidden Markov models (HMM) [2], and

conditional random fields (CRF) [3] are popular techniques for extracting aspects from the text. These methods typically rely on hand-crafted features, bigram analysis, and part-of-speech tagging to identify aspects of the text. Second, aspect polarity detection (APD) refers to identifying the semantic orientation, whether it is positive, negative, neutral, or conflicted, for example, for each aspect that is assessed inside a sentence, such as “The voice on this phone is great”. This particular review is a complimentary and positive judgment of the item, the camera. Third, aspect category detection (ACD), known as a prepared list of aspect categories, is used in the task of ACD, which aims to identify the aspect category being studied in a given statement, such as “The pizza was particularly tasty”. Food is the aspect category. It is regarded as the pizza aspect term’s hypernym. Finally, aspect category polarity (ACP) targets identifying the polarity of the sentiment of the analysed aspect categories in a given sentence. For instance, “The foods were wonderful, but the music was terrible”, is the goal of the assignment known as aspect category polarity. In this review, the author expressed a favourable, positive opinion of the cuisine category but an unfavourable, negative opinion of the ambience category.

Recently, there has been a growing interest in using transfer learning (TL) techniques for NLP tasks [4]. TL refers to the process of adapting a model pre-trained on a particular task to a new task, typically by fine-tuning the model on a new dataset. One advantage of TL is that it can significantly reduce the amount of labelled data and computational resources needed to train a model for the new task, especially when the new task is related to the original task [5,6].

The proposed model will be based on previously trained language models, which have achieved state-of-the-art performance on various NLP tasks. These models will be fine-tuned on a large annotated dataset of Arabic product reviews in order to adapt them to the tasks at hand, ATE and APD. To the best of our knowledge, there is a lack of research on the application of TL to NLP tasks in Arabic. This is particularly true for ATE and APD, which are crucial for understanding and summarizing customer reviews. Therefore, we hope to not only demonstrate the effectiveness of TL for these tasks in Arabic, but also gain insights into how the characteristics of the pre-trained models and the amount of fine-tuning data influence the performance of the TL-based model.

In particular, in this paper, we use TL techniques based on several well-performing pre-trained language models to perform ATE and APD in Arabic reviews simultaneously. The following is a summary of this paper’s core contributions:

- Upgrade the Human Annotation of Arabic Dataset (HAAD) in such a way as to make a combination of the two aspect term extraction and aspect polarity detection tasks.
- Combine fine-tuned Arabic base BERT [7] and CRF for the sake of better representation of words to solve the ATE task on an Arabic dataset. This is the first Arabic ABSA work of this kind that we are aware of.
- Utilize a cutting-edge approach for fine-tuning the BERT model to enhance the results obtained from fine-tuned ATE and APD tasks.

The rest of this paper is structured as follows. Section 2 offers a literature review of ABSA and Arabic ABSA. Section 3 offers the proposed models. Section 4 presents the specifics of the experiments and the findings of the evaluation. The paper is then concluded in Section 5, which also outlines this research’s future directions.

2. Related Work

The ABSA is a precise SA assignment that aims to extract aspects and their associated polarity from users’ audits. With Al-Smadi et al. [8]’s work, the Arabic ABSA task debuted in 2015. For each ABSA task, they also created a set of baseline models. ATE and APD both had an F1 score of 23.4%, whereas APD and ACP each received a precision of 29.7% and 42.6%.

Compared to other NLP tasks, deep learning (DL) approaches are still at an early development stage for ABSA [9] and particularly for Arabic, which is harder than, for example, English ABSA [10]. Al-Smadi et al. [11] proposed two different models using

long short-term memory (LSTM) aiming to improve a hotel's dataset results [12,13] in slots 2 and 3. The best results for slot 2 were achieved with BiLSTMCRF (FastText), a 39% improvement over baseline results (F1 value = 69.9% vs. 30.9%). On the other hand, slot 3 gave results comparable to the best model of SemEval 2016 task 5 [13] (Accuracy = 82.6%).

In Al-Dabet et al. [10]'s proposal, the model of previous opinion target expression extraction is further improved with more additional CNN layers for a character-level extraction and concatenation with word-level vectors, as well as the class's attention to grasping the main parts of the sentence. With CBOW being trained on the Wikipedia dataset, a performance reaching an F1 score of 72.8% was achieved. Various tests were also carried out both with and without the CNN model. They demonstrated that the performance of convolution tweets is positively impacted by the character-level vectors recovered by CNN.

In several downstream tasks, including SA, language models have recently produced state-of-the-art outcomes. These models can be customized for downstream NLP tasks utilizing only a little amount of labelled data after being pre-trained with a huge amount of unlabelled text. Due to this, they may overcome the dearth of annotated datasets in low-resource languages such as Arabic and spare themselves the time and resources required to build a new model from the start. OpenAI GPT [14], XLNET [15], and BERT [7] are the three most popular pre-trained language models.

To more accurately depict Arabic words, Fadel et al. [1] concatenated the embedding of BERT and Flair. AraBERT, the contextualized Arabic language model, and the embedding of Flair were combined to create the Arabic ATE model. This extended layer, BiLSTM-CRF or BiGRU-CRF, was used for sequence tagging. Both the BF-BiLSTM-CRF and the BF-BiGRU-CRF models were proposed. The experimental findings demonstrated that the suggested BF-BiLSTM-CRF configuration performed better than the baseline and other models by reaching an F1 score of 79.7% in the dataset of Arabic hotel reviews.

In Al-Smadi et al. [11]'s proposed approach, the Bi-LSTM and CRF were used by the authors at, respectively, the word and character level for AABSA. Their result outperformed the baseline on both tasks, particularly in task one, 39%, with 6% in task two. Moreover, in Al-Dabet et al. [10]'s work, the mechanism of attention for ATE was utilised, and the conducted experiments were reported to be reaching a 72.7% F1 score. Similar results were obtained using BiGRU instead of Bi-LSTM for ATE [10]. Moreover, Bensoltane and Zaki [5] proposed a BERT-BiLSTM-CRF model for ATE from the similar aforementioned news dataset.

Moreover, two DL-based methods for ABSA are proposed by Abdelgwad et al. [16] using gated recurrent unit (GRU) neural networks. The first method is a combination of bidirectional GRU, convolutional neural network (CNN), and CRF called the BGRU-CNN-CRF model, which extracts the main opinionated aspects (OTE). The second method is an interactive attention network based on bidirectional GRU (IAN-BGRU), which is used to identify the sentiment polarity towards the extracted aspects. Furthermore, Gao et al. [17] presents a short-text aspect-based sentiment analysis method using a hybrid model of CNN and BiGRU. The model takes in corpus sentences and feature words as input and outputs the emotional polarity. Furthermore, Alqaryouti et al. [18] presents a method for analysing the sentiments in reviews of smart apps that takes into account important aspects mentioned in the reviews. This approach combines the use of domain-specific vocabulary and rules to extract the relevant aspects and classify the sentiments. It employs various language processing techniques and makes use of predetermined rules and lexicons to address difficulties in sentiment analysis and produce a summary of the results.

3. Proposed Model Architecture

This section starts with a reminder of the Arabic BERT architecture and CRF. Then, we explain the proposed approach for ABSA.

3.1. BERT Models for Arabic Language

BERT is implemented using a transformer, a sizeable part that processes sequence input using an encoder and generates an expectation for the task using a decoder. Only the encoder portion of BERT is implemented in order to create a language representation model. BERT accepts a single sequence for embedding and tagging or a pair of sequences for classification as its input. Two further tokens are added at the start (CLS) and end (SEP) of the tokenized sentence before feeding to BERT. By adding a further layer (or layers) on top of BERT, and processing all of the layers at once, BERT can be improved while using fewer resources for downstream NLP tasks. The fundamental BERT architecture was trained in a variety of languages.

There are several reasons why the use of BERT would be justified in the proposed study. Firstly, BERT has been specifically designed to handle the complexities of NLP and has achieved state-of-the-art results on a wide range of NLP tasks. This makes it a strong candidate for use in the study, particularly given that the study is focused on Arabic language processing. Moreover, BERT has multiple Arabic versions that have been trained on large Arabic language corpora, which gives it the ability to capture the nuances and characteristics of the Arabic language. This is important because the specific characteristics of a language can significantly impact the performance of a language model, and using a model that has been trained on a similar language can improve its performance.

As highlighted above, there are a number of different BERT versions dedicated to the Arabic language. With 768 hidden dimensions, 12 blocks of the transformer, 12 attention heads, and a maximum sequence length of 512 tokens, Arabic BERT [19] employed the standard setup of BERT. CAMELBERT [20] was developed as a collection of BERT models that have already been pre-trained on Arabic texts of various sizes and types (modern standard Arabic (MSA), dialectal Arabic (DA), classic Arabic (CA) and a mix of the three). Similarly, the BERT base configuration was utilized by ARABERT [21]. MARBERT [22] is another large masked and pre-trained language model focused on both MSA and DA. mBERT [7] is a multilingual extension of BERT, which is trained on Wikipedia monolingual corpora in 104 languages, including Arabic.

3.2. Conditional Random Fields (CRF) Layer

Strong dependencies between labels must be taken into account in sequence-labelling tasks such as aspect extraction. While BiLSTM or BiGRU can account for long-term context information, they are unable to account for tag dependency while generating output results. These issues can be addressed through CRF. With highly interdependent output labels, the CRF layer is used. Labelling choices are collaboratively modelled with a CRF layer rather than individually, with the goal of producing the best possible global sequence of labels from an input sequence.

Conceptually, conditional random fields are an undirected graphical model for sequence labelling. CRFs can also model a much richer set of label distributions because they can define a much larger set of features, and CRFs can have arbitrary weights. Mathematically, CRF can be stated as follows. Following [1,23], we denote by $X = (x_1, x_2, \dots, x_N)$ a given input sequence, and by $Y = (y_1, y_2, \dots, y_N)$ the corresponding tag sequence.

The following is the formula for a y label sequence's conditional probability given a sequence x [24]:

$$p_{\lambda, \mu}(x|y) = \frac{1}{Z(x)} \cdot \prod_{i=1}^n \exp(\lambda_{y_{i-1}y_i} + \mu_{y_i} \cdot x_i)$$

where x is a feature vector such that $x_i \in R^m$, μ is a matrix of size $|Y| \times m$, λ is a matrix of $|Y| \times |Y|$, μ_{y_i} is the i th row in the matrix μ , and $Z(x)$ is a normalization constant known as a partition function [25]. In the equation, $\lambda_{y_{i-1}y_i}$ is a transition score that represents the score of a transition from the tag y_{i-1} to the tag y_i . The term $\mu_{y_i} \cdot x_i$ is an emission score that refers to the score of the tag y_i of the word x_i . To estimate the parameters (λ, μ) , we use

maximum likelihood estimation (MLE). The Viterbi [26] algorithm is used by the model during testing to predict the best-scoring tag sequence.

3.3. Proposed Fine-Grained Annotation of Arabic Dataset (HAAD)

The BIO scheme is widely employed for labelling words in an ATE task where B stands for the beginning of an aspect, I stands for inside an aspect, and O stands for outside an aspect, i.e., just a regular word. In this study, we used a more fine-grained annotation which jointly takes into consideration the aspect term extraction and its polarity classification (B-Positive, B-Negative, B-Conflict, B-Neutral, I-Positive, I-Negative, I-Conflict, I-Neutral, O). The biggest benefit of CRF is that it automatically picks up on some constraints for output labels that adhere to the BIO labelling scheme, which helps to validate predicted sequence labels. While learning, these restrictions are automatically learned. The following are some examples of these constraints in the context of our ATE task [1]: The opening prediction label may read 'B-Positive' or 'O', but not 'I-Positive'. The 'O I-Positive' pattern is not valid because 'I-Positive' should come before 'B-Positive'. The 'B-Negative I-Positive' pattern is invalid because of 'I-Positive', and 'I-Positive' should be preceded by 'B-Positive'.

3.4. Proposed Joint Model

In this paper, we jointly solve the ATE and APD tasks. Figure 1 shows the proposed architecture. First, the input sentence is tokenized with the associated BERT tokenizer to ensure the text is split the same way as the pre-training corpus and minimize the out-of-vocabulary terms. BERT requires input sentences of the same length with a maximum of 512 tokens. Thus, for short sentences, special [PAD] tokens are added to make sentences of equal length. BERT outputs the hidden state or the encoding vector corresponding to each token, including the special tokens. Then, these inputs are fed to a fully connected layer followed by the linear-chain CRF layer that jointly outputs the aspect and the polarity. The goal of CRF is to create a dependency between successive labels or aspects and to ensure the validity of the aspect sequence.

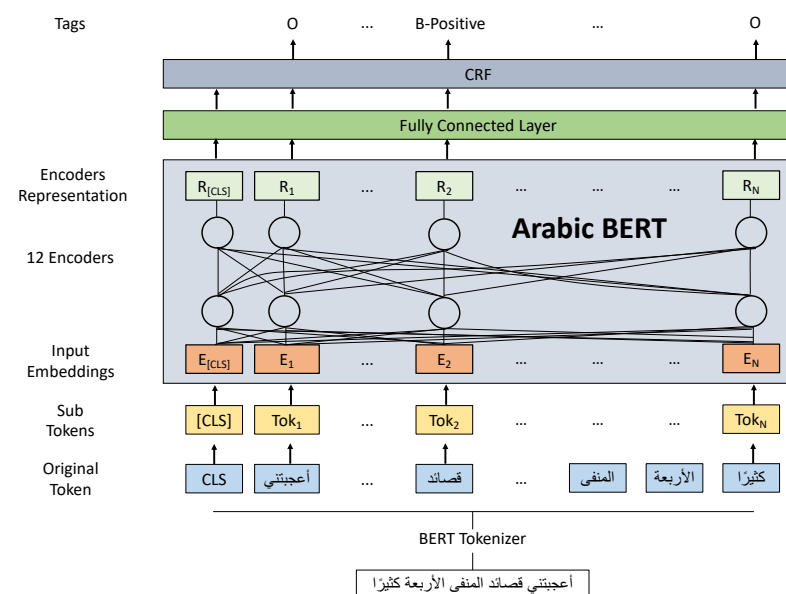


Figure 1. Architecture of Arabic BERT+CRF models [27]. An Arabic BERT-CRF model is fed text. To contextualise the word and create the final representation, tokens are employed.

4. Experiments and Results

4.1. Dataset Description

For Arabic ABSA, the HAAD dataset [8] is the main accessible dataset. The HAAD dataset contains 1513 reviews of Arabic books, each of which was described using aspect

terms, APD, ACD, and ACP. There are 2838 aspect terms in HAAD altogether. With regard to both the training and testing datasets, Table 1 presents a summary of the statistics for the sentiment polarity classes, including positive, negative, conflict, and neutral. Figure 2 shows an example of an annotated HAAD sentence.

Table 1. HAAD dataset's distribution.

Datasets	Positive	Negative	Neutral	Conflict	Total
Train	1252	855	126	26	2259
Test	124	432	22	1	579
Total	1376	1287	148	27	2838

كتاب مميز أعجبنى العديد من المقالات وبالذات في النصف الأخير من الكتاب
o o o o o o B-Positive o o o o B-Positive

Figure 2. Example of a sentence from HAAD dataset.

4.2. Hyperparameter Setting

The pre-trained Arabic BERT model [19], which had been trained on about 8.2 billion words of dialectal Arabic and MSA, was used. Particularly popular is the BERT base model, which consists of 12 attention heads, 12 covered-up layers, and a covered-up measure of 768. The model was fine-tuned for the downstream task using the Adam optimizer and the parameters listed in Table 2 with a learning rate of 1×10^{-5} , dropout rate of 0.1, hidden dropout probability of 0.3, batch size, and a variable number of epochs.

Table 2. Hyperparameter setting.

Parameter	Values/Selection
Epochs	8
Optimizer	Adam
Batch size	16
Learning rate	1×10^{-5}

4.3. Results and Discussion

In order to determine the effectiveness of the proposed approach under different BERT versions, the accuracy metric and F1 score were adopted. The experimental results are illustrated in Table 3. First, it shows that any BERT-based model can be improved by stacking a CRF layer on top of it. Second, it shows that the combination of Arabic base BERT and CRF (the proposed approach) outperforms current SOTA models in terms of accuracy, achieving a score of 95.23% and an F1 score of 47.63%.

Table 3. Performance results with different models.

BERT Model	Without CRF		With CRF	
	Accuracy (%)	F1 (%)	Accuracy (%)	F1 (%)
Baseline [8]	29.70	23.39	-	-
Arabic Base BERT	93.68	47.56	95.23	47.63
Arabic Medium BERT	93.18	30.27	93.28	35.12
Arabic Large BERT	93.00	41.65	93.76	35.50
AraBERT [21]	93.72	47.60	95.15	46.32
MarBERT [28]	92.48	40.72	92.56	39.88
CamelBERT-MSA [20]	93.05	43.97	93.89	44.10

4.4. Error Analysis

The weakness of the classifier results from both errors in the used model and errors in datasets, such as mislabelled objects. We begin by examining the limitations of the proposed models and identifying potential areas for improvement in future work.

The first error concerns the model's failure to properly handle aspect terms that have long and complex structures. Specifically, the model did not correctly identify all of the words within the aspect term, and it incorrectly assigned some words the same aspect term. For example, the following word:

الشخصية
(author's character or personality)

was wrongly tagged. Please refer to Figures 3 and 4. To overcome the issue of not tagging certain words, such as personality and book, we should augment the data sample with more sentences containing these words.

Arabic Sentence	لم	تثبت	إلا	بعض	الأراء	الشخصية	لتوفيق	الحكيم	في	بعض	القضايا
English Sentence	issues	some	on	emerged	Al-Hakim	Tawfiq	of	opinions	personal	some	Only
BIO Tagging	o	o	o	o	o	o	o	o	o	o	o

Figure 3. First type of error: long and complex structure of some aspect terms. First example.

Arabic Sentence	ربما	يكون	الكتاب	جيذا	ولكن	بروز	شخصية	الكاتب	المتعلقة	تفسد	ذلك
English Sentence	spoils	character	fawning	author's	the	but	good,	be	may	book	The
BIO Tagging	o	o	o	o	o	o	I-Negative	B-Negative	o	o	o

Figure 4. First type of error. Second example.

The second category of errors concerns the handling of implicit aspect extraction for sentiment analysis. An aspect is qualified as implicit if it is inferred from the text instead of explicitly mentioned [29]. For example, the aspect weight is implicit in the sentence “It is very light. You can carry it everywhere”.

The third category of errors is due to the use of figurative language [30] such as irony, sarcasm, and metaphor, which is considered a significant challenge in sentiment analysis. As a result of the dataset's examination, we discovered certain annotation mistakes made by human annotators with reference to the aspect terms in the HAAD test dataset. In the review id = ‘255’ (“I never liked it”), for instance, the expression “I like him” is wrongly labelled as an aspect word even though it is actually a polarity term with an implied aspect term (see Table 4).

Table 4. Third type of error: wrong label. First example.

Sentence	“I Never Liked It”
Extracted Aspect	Liked
True Aspect	Book
Aspect polarity	Negative

We also discovered a few reviews that lack human annotations, such as review id = 92, which states “Comment I will suffice to say that I still read it from time to time” (see Table 5).

Table 5. Third type of error: unlabelled sentence. Second example.

Sentence	“Comment I Will Suffice to Say That I Still Read It from Time to Time”
Extracted Aspect	-
Aspect polarity	-
Category polarity	Positive

Table 6 shows that the BERT layer combined with HMM or CRF outperforms LSTM. It also shows that it is better to stack the CRF layer on an LSTM or BERT, rather than an HMM, since the CRF has the ability to implicitly consider various constraints on the order of aspects.

Table 6. Error analysis of Arabic base BERT-CRF model

Model	Accuracy (%)
LSTM-HMM	76.85
LSTM-CRF	88.79
Arabic Base BERT-HMM	90.12
Arabic Base BERT-CRF	95.23

5. Conclusions

In this paper, we have proposed an Arabic-BERT-CRF model that combines contextual pre-trained Arabic-BERT as embedding layers and CRF as a classifier layer. The objective is to jointly solve the aspect term extraction (ATE) and aspect polarity detection (APD) tasks. The experiments were carried out using the HAAD dataset. We also enriched the dataset by manually adding polarity to each aspect. The experimental results show that the joint model Arabic-BERT-CRF outperforms the sequential model where the aspects are extracted first, then polarities are assigned.

The proposed approach was evaluated using the accuracy metric and F1 score to determine its effectiveness with different versions of BERT. The results, displayed in Table 3, demonstrate that adding a CRF layer on top of any BERT-based model can improve its performance. Additionally, the combination of Arabic base BERT and CRF (the proposed approach) achieved the highest accuracy score of 95.23% and F1 score of 47.63%, outperforming the current state-of-the-art models.

As a potential future step in this line of research, it would be interesting to explore the possibility of jointly addressing the other subtasks of Arabic aspect-based sentiment analysis. This could involve developing a model that is able to simultaneously perform multiple subtasks, such as identifying the aspect and determining the sentiment expressed towards it, rather than addressing each subtask separately. Such a model could potentially be more efficient and effective at performing the overall task of aspect-based sentiment analysis. Additionally, studying the feasibility of jointly solving the subtasks could provide insights into the relationships and dependencies between the subtasks, and potentially lead to the development of new approaches for addressing the overall task.

Author Contributions: Conceptualization, M.A.; methodology, H.C., F.J.; software, H.C., F.J.; validation, H.C., F.J.; formal analysis, M.A.; investigation, H.C.; resources, H.C.; data curation, H.C.; writing—original draft preparation, H.C.; writing—review and editing, M.A., F.J.; visualization, M.A.; supervision, F.J.; project administration, F.J. All authors have read and agreed to the published version of the manuscript.

Funding: The researchers would like to thank the Deanship of Scientific Research, Qassim University for funding the publication of this project.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Fadel, A.S.; Saleh, M.E.; Abulnaja, O.A. Arabic Aspect Extraction based on Stacked Contextualized Embedding with Deep Learning. *IEEE Access* **2022**, *10*, 30526–30535. [\[CrossRef\]](#)
2. Jin, W.; Ho, H.H.; Srihari, R.K. A novel lexicalized HMM-based learning framework for web opinion mining. In Proceedings of the 26th Annual International Conference on Machine Learning, Montreal, QC, Canada, 14–18 June 2009; ACM: New York, NY, USA, 2009; pp. 465–472.
3. Jakob, N.; Gurevych, I. Extracting opinion targets in a single and cross-domain setting with conditional random fields. In Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, Cambridge, MA, USA, 9–11 October 2010; pp. 1035–1045.
4. Mozafari, M.; Farahbakhsh, R.; Crespi, N. A BERT-based transfer learning approach for hate speech detection in online social media. In *International Conference on Complex Networks and Their Applications*; Springer: Cham, Switzerland, 2019; pp. 928–940.
5. Bensoltane, R.; Zaki, T. Towards Arabic aspect-based sentiment analysis: A transfer learning-based approach. *Soc. Netw. Anal. Min.* **2022**, *12*, 7. [\[CrossRef\]](#)
6. Chouikhi, H.; Alsuhaibani, M. Deep Transformer Language Models for Arabic Text Summarization: A Comparison Study. *Appl. Sci.* **2022**, *12*, 11944. [\[CrossRef\]](#)
7. Kenton, J.D.M.W.C.; Toutanova, L.K. Bert: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the NAACL-HLT, Minneapolis, MN, USA, 2–7 June 2019; pp. 4171–4186.
8. Al-Smadi, M.; Qawasmeh, O.; Talafha, B.; Quwaider, M. Human annotated arabic dataset of book reviews for aspect based sentiment analysis. In Proceedings of the 2015 3rd International Conference on Future Internet of Things and Cloud, Rome, Italy, 24–26 August 2015; pp. 726–730.
9. Oueslati, O.; Cambria, E.; HajHmida, M.B.; Ounelli, H. A review of sentiment analysis research in Arabic language. *Future Gener. Comput. Syst.* **2020**, *112*, 408–430. [\[CrossRef\]](#)
10. Al-Dabet, S.; Tedmori, S.; Al-Smadi, M. Extracting opinion targets using attention-based neural model. *SN Comput. Sci.* **2020**, *1*, 242. [\[CrossRef\]](#)
11. Al-Smadi, M.; Talafha, B.; Al-Ayyoub, M.; Jararweh, Y. Using long short-term memory deep neural networks for aspect-based sentiment analysis of Arabic reviews. *Int. J. Mach. Learn. Cybern.* **2019**, *10*, 2163–2175. [\[CrossRef\]](#)
12. Mohammad, A.S.; Qwasmeh, O.; Talafha, B.; Al-Ayyoub, M.; Jararweh, Y.; Benkhelifa, E. An enhanced framework for aspect-based sentiment analysis of Hotels’ reviews: Arabic reviews case study. In Proceedings of the 2016 11th International Conference for Internet Technology and Secured Transactions (ICITST), Barcelona, Spain, 5–7 December 2016; pp. 98–103.
13. Pontiki, M.; Galanis, D.; Papageorgiou, H.; Androustopoulos, I.; Manandhar, S.; Al-Smadi, M.; Al-Ayyoub, M.; Zhao, Y.; Qin, B.; De Clercq, O.; et al. Semeval-2016 task 5: Aspect based sentiment analysis. In Proceedings of the International Workshop on Semantic Evaluation, San Diego, CA, USA, 16–17 June 2016; pp. 19–30.
14. Radford, A.; Wu, J.; Child, R.; Luan, D.; Amodei, D.; Sutskever, I. Language models are unsupervised multitask learners. *OpenAI Blog* **2019**, *1*, 9.
15. Yang, Z.; Dai, Z.; Yang, Y.; Carbonell, J.; Salakhutdinov, R.R.; Le, Q.V. Xlnet: Generalized autoregressive pretraining for language understanding. *Adv. Neural Inf. Process. Syst.* **2019**, *32*, 5753–5763.
16. Abdelgwad, M.M.; Soliman, T.H.A.; Taloba, A.I.; Farghaly, M.F. Arabic aspect based sentiment analysis using bidirectional GRU based models. *J. King Saud Univ.-Comput. Inf. Sci.* **2022**, *34*, 6652–6662.
17. Gao, Z.; Li, Z.; Luo, J.; Li, X. Short Text Aspect-Based Sentiment Analysis Based on CNN+ BiGRU. *Appl. Sci.* **2022**, *12*, 2707. [\[CrossRef\]](#)
18. Alqaryouti, O.; Siyam, N.; Monem, A.A.; Shaalan, K. Aspect-based sentiment analysis using smart government review data. *Appl. Comput. Inform.* **2020**, *16*, 1–20. [\[CrossRef\]](#)
19. Safaya, A.; Abdullatif, M.; Yuret, D. Kuisail at semeval-2020 task 12: Bert-cnn for offensive speech identification in social media. In Proceedings of the Fourteenth Workshop on Semantic Evaluation, Barcelona, Spain, 12–13 December 2020; pp. 2054–2059.
20. Inoue, G.; Alhafni, B.; Baimukan, N.; Bouamor, H.; Habash, N. The interplay of variant, size, and task type in Arabic pre-trained language models. *arXiv* **2021**, arXiv:2103.06678.
21. Antoun, W.; Baly, F.; Hajj, H. Arabert: Transformer-based model for arabic language understanding. *arXiv* **2020**, arXiv:2003.00104.
22. Abdul-Mageed, M.; Elmadany, A.; Nagoudi, E.M.B. ARBERT & MARBERT: Deep bidirectional transformers for Arabic. *arXiv* **2020**, arXiv:2101.01785.
23. Chen, Q.; Zeng, X.; Zhu, J.; Zhang, Y.; Lin, B.; Yang, Y.; Jiang, D. Rethinking the Value of Gazetteer in Chinese Named Entity Recognition. In *CCF International Conference on Natural Language Processing and Chinese Computing*; Springer: Cham, Switzerland, 2022; pp. 285–297.

24. Lafferty, J.; McCallum, A.; Pereira, F.C. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In Proceedings of the 18th International Conference on Machine Learning 2001 (ICML 2001), Williamstown, MA, USA, 28 June–1 July 2001.
25. Tutubalina, E.; Nikolenko, S. Combination of deep recurrent neural networks and conditional random fields for extracting adverse drug reactions from user reviews. *J. Healthc. Eng.* **2017**, *2017*, 9451342. [[CrossRef](#)] [[PubMed](#)]
26. Forney, G.D. The viterbi algorithm. *Proc. IEEE* **1973**, *61*, 268–278. [[CrossRef](#)]
27. Al-Qurishi, M.S.; Souissi, R. Arabic Named Entity Recognition Using Transformer-based-CRF Model. In Proceedings of the Fourth International Conference on Natural Language and Speech Processing (ICNLSP 2021), Trento, Italy, 12–13 November 2021; pp. 262–271.
28. Abdul-Mageed, M.; Elmadany, A. ARBERT & MARBERT: Deep Bidirectional Transformers for Arabic. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), Virtual, 1–6 August 2021; pp. 7088–7105.
29. Ganganwar, V.; Rajalakshmi, R. Implicit aspect extraction for sentiment analysis: A survey of recent approaches. *Procedia Comput. Sci.* **2019**, *165*, 485–491. [[CrossRef](#)]
30. Ghosh, A.; Li, G.; Veale, T.; Rosso, P.; Shutova, E.; Barnden, J.; Reyes, A. Semeval-2015 task 11: Sentiment analysis of figurative language in twitter. In Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), Denver, CO, USA, 4–5 June 2015; pp. 470–478.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.