ACIT'2017
The International Arab Conference on Information Technology
Yassmine Hammamet, Tunisia
December 22-24, 2017

Aspect-Based Sentiment Analysis of Arabic Laptop Reviews

Mahmoud Al-Ayyoub,[1] Amal Gigieh,[2] Areej Al-Qwaqenah,[2] Mohammed N. Al-Kabi,[3] Bashar Talafhah,[1] and Izzat Alsmadi[4]
[1] Computer Science Department, Jordan University of Science and Technology, Irbid, Jordan
E-mails: maalshbool@just.edu.jo, talafha@live.com
[2] Computer Department, Ajloun College, Al-Balqa' Applied University, Ajloun, Jordan
E-mail: amal.qiqi@bau.edu.jo, areej.cis@bau.edu.jo
[3] Information Technology Department, Al-Buraimi University College, Buraimi, Oman
E-mail: mohammed@buc.edu.om
[4] Department of Computing and Cyber Security, Texas A&M University-San Antonio, San Antonio, TX 78224, USA
E-mail: ialsmadi@tamusa.edu

Abstract: *Sentiment Analysis (SA) is one of the hottest research areas in Natural Language Processing (NLP) with vast commercial as well as academic applications. One of the most interesting versions of SA is called Aspect-Based SA (ABSA). Currently, most of the researchers focus on English text. Other languages such as Arabic have received less attention. To the best of our knowledge, only few papers have addressed ABSA of Arabic reviews and they have all been applied on only three datasets. In this work, we demonstrate our efforts to build the Arabic Laptops Reviews (ALR) dataset, which focuses on laptops reviews written in Arabic. To make it easy to use, the ALR dataset is prepared according to the annotation scheme of SemEval16-Task5. The annotation scheme considers two problems: aspect category prediction and sentiment polarity label prediction. It also comes with an evaluation procedure that extracts n-grams' features and employs a Support Vector Machine (SVM) classifier in order to allow researchers to gauge and compare the performance of their systems. The evaluation results show that there is a lot of room for improvements in the performance of the SVM classifier for the aspect category prediction problem. As for the sentiment polarity label prediction, SVM's accuracy is actually high.*

Keywords: *Aspect Based Sentiment Analysis; Arabic Laptop Reviews; Aspect Category Prediction; Sentiment Polarity Label Prediction; N-Gram Features; SVM Classifier;*

1. Introduction

The explosive growth of User Generated Contents (UGC) represents a wealth of raw data that is of interest to many parties in the industry as well as academia. One of the hottest areas of UGC analysis is Sentiment Analysis (SA) or Opinion Mining (OM), which is concerned with extracting the sentiments conveyed in a piece of Text/Image/Audio/Video. Over the past decade, SA has grown and become more complicated in order to meet more specific demands (especially, those from the commercial applications). This has made SA one of the major problems in Natural Language Processing (NLP) and Data Mining [1].

The goal of typical SA systems is to determine a single sentiment polarity of each opinionated sentence or review. This is not always practical or useful as sentences or reviews might contain several opinions about different aspects of a certain entity. Moreover, these opinions might be conflicting/contradicting with each other. For example, in a single sentence, the reviewer might praise the performance of the graphics card in a laptop/Desktop computer and criticize the lifetime of its battery. Hence, a finer grained analysis of such reviews is required, which gave rise to feature-based or aspect-based SA (ABSA) [2].

ABSA is an extension of SA which is concerned with extracting all opinions in a sentence or a review along with the entities and aspects they are targeting and their sentiment polarity values. ABSA is very useful, for examples such as the one discussed in the previous paragraph. Due to its importance, ABSA, has been the focus of attention at high profile NLP conferences and workshops such as SemEval [3-5].

Semantic Evaluation (SemEval) is one of the most prestigious workshops in NLP. It is held annually and offers the scientific community with a set of exercises to evaluate SA systems. In 2014, SemEval hosted the first shared task on ABSA [3]. The task provided the scientific community with benchmark datasets as well as common evaluation procedures. Interested researchers can download the datasets and use them to train their systems. Then, they can submit their systems for evaluation after which the winning teams are announced.

The success of this task has led to repeated tasks in the following two years [4, 5], where the task grew in its coverage of multiple domains and multiple languages and the challenges it poses. In fact, the latest ABSA task hosted by SemEval provided 19 training and 20 testing datasets for eight languages and seven domains. Moreover, the baseline evaluation procedure employed a classifier that is known to perform well for NLP tasks, which is the Support Vector Machine (SVM) classifier. This has added extra challenges to the participating teams. The task

was very popular that it attracted 245 submissions from 29 teams.

Unfortunately, most of the previous discussion pertains mainly to the English language. The work on ABSA of other languages is very limited. The Arabic language is no exception. To the best of our knowledge, not many papers have been published on ABSA of Arabic text [5-19]. Some of these papers [5, 13-19] are about the 2016 version of the SemEval ABSA shared task which provided an Arabic dataset of hotel reviews. In addition to baseline classification performed on this dataset [1], this dataset attracted three submissions which represented cross-lingual efforts [13-16]. The other works consider either a dataset of book reviews [6-8] or a dataset of news posts and their comments [9-12]. In this work, we present the Arabic Laptop Reviews (ALR) dataset, which is annotated according the SemEval16-Task5 guidelines. Considering laptop reviews has obvious commercial impact which made it one of seven domains of SemEval16-Task5 [5].

The remainder of this paper is organized as follows. The next section discusses the papers that are most relevant to our work. Section 3 presents the different ABSA tasks considered in this work. Section 4 introduces our ALR dataset and shows a preliminary discussion of the collected corpus as well as the results of the analysis performed on it. Section 5 presents concluding remarks about this paper and discussion of future plans to extend this corpus.

2. Related works

SA has been well investigated over the past two decades. This is evident in the high numbers of papers published on it, events dedicated to it and systems developed based on its ideas [20-25]. However, there are still many challenges associated with SA. One of these challenges is the ability to perform fine grained analysis of reviews with multiple targets and conflicting sentiments. This gave rise to the important problem of ABSA [2].

Most of the existing works on ABSA have focused on the English language with little attention paid to other languages such as Arabic. One of the notable exceptions to this is the recent effort by the organizers of the ABSA shared task hosted by SemEval16, who expanded their coverage to include eight languages including Arabic [5]. In this work, we focus on the problem of ABSA of Arabic reviews. Discussion of the state-of-the-art in ABSA of English reviews is beyond the scope of this paper. Interested readers are referred to [2-5, 26-30].

Despite its importance, the field of Arabic NLP (ANLP) has a very limited set of resources compared to the other languages [31, 32]. Being part of ANLP, Arabic SA suffers from the same problem. There have been some decent efforts to provide resources such as publicly available standard benchmark datasets for Arabic SA [33-39]. However, the efforts towards Arabic ABSA have been limited.

The start was with the Human Annotated Arabic Dataset (HAAD) of book reviews [6]. The authors relied on a publicly available dataset of more than 63,000 Arabic book reviews called the Large Arabic Book Reviews (LABR) dataset [40], which has been the focus of many papers [41-43]. They selected a small number of reviews and re-annotated them for ABSA purposes. The HAAD dataset has been studied

Arabic Review

اول شي الاتش بي يحتر بسرعه وادائه تحت المتوسط وغير جيد للالعاب وبرامج التصميم يعني بشل عام غير جيد للاستعمال الثقيل ونوعيه كرت الشاشه ماله صح ٢ جي بي بس هذي نوعيه ضعيفه جدا وعندك الهاردسك صغير شوي والمعالج اي٥ الحين ما راح يفيدك غير اي٧ والمعالج شكله خماسي الخيوط من الجيل الثاني ورباعي النواه هذا جيد نوعا ما لكن لا انصحك به للاستفسار 96523275

English translation

The first thing is that HP gets hot quickly. Its performance is below average. It is not good for games and design software. Hence, in general, it is not good for heavy usage. The type of the graphics card is incorrect. Even though it has 2GB, but the type is very weak. You have a hard disk that is rather small. The processor is i5; nowadays, nothing is useful but i7. The processor seems like quintuple-threaded second-generation quad-code. This is kind of good, but I don't recommend it. For inquiry, 96523275.


```
<Review rid="bbcode_postedby7542">
 <sentences>
  <sentence id="bbcode_postedby7542:0">
   <text>اول شي الاتش بي يحتر بسرعه وادائه تحت المتوسط</text>
   <Opinions>
    <Opinion category="FANS_COOLING#QUALITY" polarity="negative"/>
    <Opinion category="LAPTOP#OPERATION_PERFORMANCE" polarity="negative"/>
   </Opinions>
  </sentence>
  <sentence id="bbcode_postedby7542:1">
   <text>
    وغير جيد للالعاب وبرامج التصميم يعني بشل عام غير جيد للاستعمال الثقيل
   </text>
   <Opinions>
    <Opinion category="LAPTOP#OPERATION_PERFORMANCE" polarity="negative"/>
    <Opinion category="LAPTOP#USABILITY" polarity="negative"/>
   </Opinions>
  </sentence>
  <sentence id="bbcode_postedby7542:2">
   <text>
    ونوعيه كرت الشاشه ماله صح 2 جي بي بس هذي نوعيه ضعيفه جدا وعندك الهاردسك صغير شوي
   </text>
   <Opinions>
    <Opinion category="GRAPHICS#QUALITY" polarity="negative"/>
    <Opinion category="HARDWARE#QUALITY" polarity="negative"/>
    <Opinion category="OPTICAL_DRIVES#QUALITY" polarity="negative"/>
   </Opinions>
  </sentence>
  <sentence id="bbcode_postedby7542:3">
   <text>
    والمعالج اي5 الحين ما راح يفيدك غير اي7 والمعالج شكله خماسي الخيوط من الجيل الثاني ورباعي النواه
   </text>
   <Opinions>
    <Opinion category="CPU#OPERATION_PERFORMANCE" polarity="negative"/>
    <Opinion category="CPU#DESIGN_FEATURES" polarity="negative"/>
   </Opinions>
  </sentence>
  <sentence id="bbcode_postedby7542:4">
   <text>96523275 هذا جيد نوعا ما لكن لا انصحك به للاستفسار</text>
   <Opinions>
    <Opinion category="LAPTOP#GENERAL" polarity="positive"/>
   </Opinions>
  </sentence>
 </sentences>
</Review>
```


Figure 1. The left side of the figure shows a sample laptop review in Arabic along with its translation. The right side shows the XML formatted annotation of each sentence of this review.

in later papers by the same group [7, 8].

The second Arabic dataset annotated for ABSA was a political one. In [9-12], the authors discussed their efforts to collect news posts and comments from Facebook about the 2014 Gaza Attacks for the purpose of evaluating the news affect on readers. They used an ABSA approach, hence, their dataset was annotated for ABSA purposes.

The latest dataset of interest is one of the datasets of SemEval16-Task5 [5]. It consists of hotel reviews annotated for ABSA. In addition to the provided baseline classifier, this dataset attracted the attention of many groups working on cross-lingual ABSA [13-16] in addition to efforts focusing on Arabic text [17-19].

Finally, there have been other efforts related to ABSA such as [44, 45]. These efforts focused on extracting opinion targets, which is one of the ABSA tasks as discussed in the following section.

3. ABSA Tasks

Three subtasks appeared in the SemEval16-Task5 [5]: sentence-level ABSA, text-level ABSA and out-of-domain ABSA. We limit our attention in this paper to the first two subtasks and leave the third one for future extensions of this work.

Subtask 1 (SB1): Sentence-level ABSA. As the name suggests, this subtask deals with sentences. So, given an opinionated sentence about a target entity, the goal is to identify the targets and polarities of all opinions in that sentence. The target of each opinion is expressed by an Entity#Attribute (E#A) pair, where the E and A are selected from a predefined list of [5]. As for the polarity classes considered, they are: positive, negative and/or neutral.

Subtask 2 (SB2): Text-level ABSA. This subtask is similar to the first one except that it deals with the entire review instead of handling each sentence separately; Given an opinionated review about a target entity, the goal is to identify the targets (represented as E#A pairs) and polarities (positive, negative or neutral) of all opinions in the review.

An example of a review is shown in Figure 1. The figure highlights the E#A pair and polarity of each opinion.

4. ALR Dataset

This section is dedicated to discussing the Arabic Laptops Reviews (ALR) dataset. We start by discussing the collection and annotation processes before discussing the annotation scheme we follow.

4.1. Collection and Annotation

At this stage, the reviews of the ALR dataset are mainly about laptops. They are collected from several websites including souq.com (an English-Arabic online shopping website that has been described as the Amazon of the Middle East)[1] as well as a number of web forums. The reviews are carefully selected to avoid having spam or irrelevant reviews, which require special attentions [46]. The reviews are annotated for ABSA purposes by identifying the targets and polarities of all opinions (see Figure 1). The selection and annotation are performed by two Arabic native speakers.

The ALR dataset consists of 3052 opinions taken from 1028 reviews. These opinions are unevenly distributed across the different pairs of E#A under consideration. In fact, about one sixth of the opinions are general ones. Moreover, good percentages of the opinions are targeted towards design features and price of the laptops in general. On the other hand, detailed reviews about specific components of the laptops are neither common nor comprehensive. In most of the time, if the users want to provide details in their reviews about the physical components of the laptops, they would discuss the keyboard, the battery or the display of the laptop. As for the software, they would rarely discuss anything other than the operating system. Other important aspects that are not commonly discussed in the reviews in our dataset include graphics card, cooling, support and warranty. These observations reflect the rather shallow nature of the reviews, which is one of the major challenges we observed while collecting the reviews. In fact, scrolling through the Arabic reviews of laptops reveals that most of them are either spam or very short and general.

Table 1. Basic statistics of the ALR dataset

Number of Reviews	1028
Number of Sentences	1753
Number of Opinions	3052
Opinions with Polarity	3052
Positive Opinions	2004
Negative Opinions	1048
Category Values	131

Table 1 shows the basic statistics related to ALR dataset. It can be observed from the table that the reviews are short as single-sentence reviews are very common. The table also shows that the majority of the opinions are positive.

Table 2. Word-level statistics of the ALR dataset.

Opinion	Total	Average No. of Words	Min Word	Max Word
Negative	1048	15.60	1	163
Positive	2004	11.43	1	65

[1] http://www.inc.com/laura-montini/meet-souq-the-e-commerce-giant-of-the-middle-east.html

All	3052	8.5	1	163

Table 2 presents some word-level statistics of the ALR dataset. From the table, it can be seen that the average size of negative opinions is larger than that of the positive ones. The justification is simple. We have observed that, in many cases, a reviewer with a negative opinion tend to include a justification to his/her point of view. The average size of neutral opinions is the largest, but one may argue that the number of neutral opinions in this dataset is relatively small when compared with negative and positive opinions.

Table 3 presents the data related to the ALR's opinions measured in terms of characters. As expected the results of Table 3 is similar of those presented in previous Table 2, where Table 2 uses word measure while Table 3 uses character measure.

Table 3. Character-level statistics of the ALR dataset

Opinion	Total	Average No. of Chars	Min Char	Max Char
Negative	1048	84.30	4	896
Positive	2004	63.63	5	367
All	3052	73.96	4	896

The 3052 opinions of our dataset are categorized into 131 different categories. Table 4 shows the top 20 categories with their frequencies in the ALR dataset. These values reflect the main concerns of laptop users that include the design, price, performance, quality, laptop vendor, usability, portability, keyboard, battery, brand name of the laptop vendor, monitor features, the quality of the monitor, quality of cooling fans, Operating System, CPU performance, Shipping, and graphics cards features and performances.

Table 4. The top 20 categories with their frequencies in the ALR dataset

Category Name	Frequency
LAPTOP#GENERAL	506
LAPTOP#DESIGN_FEATURES	465
LAPTOP#PRICE	397
LAPTOP#OPERATION_PERFORMANCE	221
LAPTOP#QUALITY	118
COMPANY#GENERAL	92
LAPTOP#USABILITY	82
LAPTOP#PORTABILITY	81
KEYBOARD#DESIGN_FEATURES	63
BATTERY#QUALITY	54
COMPANY#QUALITY	49
DISPLAY#DESIGN_FEATURES	49
DISPLAY#QUALITY	48
FANS_COOLING#QUALITY	48
GRAPHICS#QUALITY	39
CPU#OPERATION_PERFORMANCE	35
OS#MISCELLANEOUS	34
SHIPPING#GENERAL	26
GRAPHICS#DESIGN_FEATURES	23
GRAPHICS#OPERATION_PERFORMANCE	23

4.2. Annotation Scheme

Despite its importance, ABSA has yet to receive the proper attention it deserves within the Arabic NLP and data mining communities. This is manifested in the very small number of available datasets for ABSA of Arabic reviews. As mentioned earlier, only three such datasets are available to the best of our knowledge [1, 5, 10]. All of them follow the SemEval schemes or something similar. E.g., Hu and Liu [47] presented one of the earliest datasets for feature-based SA where the reviews are presented in a simple scheme to be classified into negative/positive (e.g., Digital Camera, Feature: picture quality, 250 positive reviews, 35 negative reviews). In a later work, the authors of [48] presented their restaurants reviews dataset annotated using category-based annotation on the sentence-level. They considered six categories such as Price and Food and four polarity labels (Positive, Negative, Neutral and Conflict).

Among the most common schemes for ABSA are the ones used by the SemEval ABSA shared tasks [3-5]. Despite having certain differences among them, the three versions of the SemEval ABSA tasks provide benchmark datasets that are well-formatted and stored in XML files. To ensure that our dataset is usable by as many researchers as possible, we follow the SemEval16-Task5 annotation scheme.

An example of a review along with its XML-formatted annotation is depicted in Figure 2. As shown in the figure, each opinion in our dataset is annotated using the following XML element:

<Opinion category=" " polarity=" "/>

5. Evaluation

The ALR dataset is provided with a baseline classifier. The classifier is based on [4, 5] and it is discussed in this section. However, before discussing the classifier, we briefly discuss the evaluation metrics.

5.1. Evaluation Metrics

In this work, we consider two prediction problems: aspect category prediction and sentiment polarity label prediction. For the former problem, the main performance measure we consider is the F_1 measure. It is computed by taking the harmonic mean of the Precision (P) and Recall (R) as follows.

$$F_1 = \frac{2 \times P \times R}{P + R} \quad (1)$$

where the Precision (P) and Recall (R) are computed according to the following equations.

$$P = \frac{|S \cap G|}{|S|} \quad (2)$$

$$R = \frac{|S \cap G|}{|G|} \quad (3)$$

where S and G are the predicted and gold aspect categories, respectively.

5.2. Evaluation Process

Before discussing the classification process, it is worth mentioning that the testing/evaluation technique we follow is the hold-out method. We make this choice based on the SemEval schemes.

Similar to SemEval16-Task5 [5], the evaluation is executed in two phases. In the first phase; Phase A, the system is evaluated based on the returned aspect categories for SB1. As for SB2, the respective text-level categories had to be identified. In the second phase; Phase B, the gold annotations for the test sets of Phase A are provided and the system is evaluated based on the returned sentiment polarity values. Similar to SemEval15-Task12 [4], the F_1 scores are calculated for the aspect categories by comparing the returned annotations with the gold annotations (using micro-averaging). For evaluating aspect categories, duplicate occurrences of categories are ignored in both SB1 and SB2. In Phase B, the sentiment polarity classification is evaluated by calculating the accuracy defined as the number of correctly predicted polarity labels of the (gold) aspect categories, divided by the total number of the gold aspect categories. We briefly discuss the baseline classifiers in the following paragraphs.

Aspect Categories: For the category (E#A) extraction problem of SB1, an SVM classifier with a linear kernel is trained. In particular, the classifier extracts *n* unigram features from each sentence in the training subset. The category value (e.g., "LAPTOP#PRICE") of the opinion is used as the correct label of the feature vector. Similarly, for each test sentence *s*, a feature vector is built and the trained SVM is used to predict the probabilities of assigning each possible category to *s* (e.g., {"SERVICE#GENERAL", 0.2}, {"RESTAURANT#GENERAL", 0.4}. Then, a threshold *t* (in our experiments, it is set to 0.2) is used to decide which of the categories will be assigned to *s*. As features, the 1,000 most frequent unigrams of the training subset are used after excluding stop-words. As for SB2, the sentence-level opinions returned by the SB1 baseline are copied to the text level and duplicates are removed.

Sentiment Polarities: For the polarity prediction problem of SB1, an SVM classifier with a linear kernel is trained. As in aspect categories prediction, *n* unigram features are extracted from each sentence in the training subset. In addition, an integer-valued feature that indicates the category of the opinion is used. The correct label for the extracted training feature vector is the corresponding polarity value (e.g., "positive"). Then, for each category of a test sentence *s*, a feature vector is built and classified using the trained SVM.

As for SB2, for each text-level aspect category *c*, the baseline traverses the predicted sentence-level tuples of the same category returned by the respective SB1 baseline and counts the polarity labels (positive, negative, neutral). Finally, the polarity label with the highest frequency is assigned to the text-level category *c*. If there are no sentence-level tuples for the same *c*, the polarity label is determined based on all tuples regardless of *c*.

The baseline systems and evaluation scripts are implemented in Java. The LibSVM package [49] is used for SVM training and prediction.

5.3. Results

For the category prediction problem, the precision is 0.529, the recall is 0.225, and the F_1 measure is 0.315. This is a low value for such a critical task. However, the justification might be in the small size of the dataset and the highly imbalanced distribution of the reviews across the different categories under consideration. Another issue is that fact that the reviews are written different versions of colloquial Arabic. The low results also highlight the fact that the problem of predicting E#A pairs of laptop reviews is complex and requires a large and comprehensive training set.

As for the sentiment polarity label prediction problem, the results are much better. The SVM classifier is able to correctly predict 218 sentiment polarities out of the 298 testing ones. This is an accuracy of 0.732, which is quite significant given the small size of the dataset.

6. Conclusion and Future Work

In this work, we discuss our efforts to build the Arabic Laptops Reviews (ALR) dataset, to focus on laptops reviews written in Arabic. To make it easier to use, the ALR dataset is prepared according to the annotation scheme of SemEval16-Task5. The annotation scheme considers two problems: aspect category prediction and sentiment polarity label prediction. It also comes with an evaluation procedure that extracts *n*-grams' features and employs an SVM classifier in order to allow researchers to compare the performance of their systems. The evaluation results show that there is a lot of room for improvement in the performance of the SVM classifier for the aspect category prediction problem. As for the sentiment polarity label prediction, SVM's accuracy is acceptable.

There are many possible future directions of this work. The most relevant one is to expand the dataset and make it more comprehensive. Adding more reviews about other products similar to laptops such as smart phones is also an interesting extension. After expanding the dataset, we plan to improve the feature extraction and classification steps. Experimenting with a classifier such as Conditional Random Field (CRF) or any of the popular deep learning techniques such as Recurrent Neural Networks (RNN) seems appealing.

References

[1] Liu, Bing. "Sentiment analysis and opinion mining." Synthesis lectures on human language technologies 5.1 (2012): 1-167.

[2] Pavlopoulos, Ioannis. "Aspect based sentiment analysis." Athens University of Economics and Business (2014).

[3] Pontiki, Maria, et al. "SemEval-2014 Task 4: Aspect Based Sentiment Analysis." Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), Association for Computational Linguistics, Dublin, Ireland. 2014.

[4] Pontiki, Maria, et al. "SemEval-2015 Task 12: Aspect Based Sentiment Analysis." Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), Association for Computational Linguistics, Denver, Colorado. 2015.

[5] Pontiki, Maria, et al. "SemEval-2016 Task 5: Aspect Based Sentiment Analysis." Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval 2016), Association for Computational Linguistics, San Diego, California. 2016.

[6] Al-Smadi, Mohammad, et al. "Human annotated Arabic dataset of book reviews for aspect based sentiment analysis." Future Internet of Things and Cloud (FiCloud), 2015 3rd International Conference on. IEEE, 2015.

[7] Obaidat, Islam, et al. "Enhancing the determination of aspect categories and their polarities in arabic reviews using lexicon-based approaches." Applied Electrical Engineering and Computing Technologies (AEECT), 2015 IEEE Jordan Conference on. IEEE, 2015.

[8] Al-Smadi, Mohammad, et al. "Using Enhanced Lexicon-Based Approaches for the Determination of Aspect Categories and Their Polarities in Arabic Reviews." International Journal of Information Technology and Web Engineering (IJITWE) 11.3 (2016): 15-31.

[9] Al-Smadi, Mohammad, et al. "Using aspect-based sentiment analysis to evaluate arabic news affect on readers." 2015 IEEE/ACM 8th International Conference on Utility and Cloud Computing (UCC). IEEE, 2015.

[10] Al-Smadi, Mohammad, et al. "An Aspect-Based Sentiment Analysis Approach to Evaluating Arabic News Affect on Readers." Journal of Universal Computer Science 22 (2016): 630-649.

[11] Al-Sarhan, Huda, et al. "Framework for affective news analysis of Arabic news: 2014 Gaza attacks case study." 2016 7th International Conference on Information and Communication Systems (ICICS). IEEE, 2016.

[12] Al-Ayyoub, Mahmoud, et al. "Framework for Affective News Analysis of Arabic News: 2014 Gaza Attacks Case Study." j jucs 23 (2017): 327-352.

[13] Kumar, Ayush, et al. "IIT-TUDA at SemEval-2016 Task 5: Beyond Sentiment Lexicon: Combining Domain Dependency and Distributional Semantics Features for Aspect Based Sentiment Analysis."

[14] Ruder, Sebastian, Parsa Ghaffari, and John G. Breslin. "NSIGHT-1 at SemEval-2016 Task 5: Deep Learning for Multilingual Aspect-based Sentiment Analysis." arXiv preprint arXiv:1609.02748 (2016).

[15] Ruder, Sebastian, Parsa Ghaffari, and John G. Breslin. "A Hierarchical Model of Reviews for Aspect-based Sentiment Analysis." arXiv preprint arXiv:1609.02745 (2016).

[16] Tamchyna, Aleš, and Kateřina Veselovská. "UFAL at SemEval-2016 Task 5: Recurrent Neural Networks for Sentence Classification." Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016). 2016.

[17] AL-Smadi, Mohammad, et al. "An enhanced framework for aspect-based sentiment analysis of Hotels' reviews: Arabic reviews case study." Internet Technology and Secured Transactions (ICITST), 2016 11th International Conference for. IEEE, 2016.

[18] AL-Smadi, Mohammad, et al. "Deep Recurrent Neural Network vs. Support Vector Machine for Aspect-Based Sentiment Analysis of Arabic Hotels' Reviews." Journal of Computational Science (2017).

[19] Alawami, Alawya. Aspect extraction for sentiment analysis in Arabic dialect. Diss. University of Pittsburgh, 2017.

[20] Feldman, Ronen. "Techniques and applications for sentiment analysis." Communications of the ACM 56.4 (2013): 82-89.

[21] AlKhatib, Khalid, et al. "On the use of arabic tweets to predict stock market changes in the arab world." International Journal of Advanced Computer Science and Applications (IJACSA)7.5 (2016): 560-566.

[22] Medhat, Walaa, Ahmed Hassan, and Hoda Korashy. "Sentiment analysis algorithms and applications: A survey." Ain Shams Engineering Journal 5.4 (2014): 1093-1113.

[23] Rabab'ah, Abdullateef, et al. "Measuring the controversy level of arabic trending topics on twitter." Information and Communication Systems

(ICICS), 2016 7th International Conference on. IEEE, 2016.

[24] Al-Ayyoub, Mahmoud, et al. "Studying the controversy in online crowds' interactions." Applied Soft Computing (2017).

[25] Zimbra, David, Manoochehr Ghiassi, and Sean Lee. "Brand-related Twitter sentiment analysis using feature engineering and the dynamic architecture for artificial neural networks." System Sciences (HICSS), 2016 49th Hawaii International Conference on. IEEE, 2016.

[26] Wang, Bo, and Min Liu. "Deep Learning for Aspect-Based Sentiment Analysis." (2015): 1-9.

[27] Zhang, Lei, and Bing Liu. "Aspect and entity extraction for opinion mining. "Data mining and knowledge discovery for big data. Springer Berlin Heidelberg, 2014. 1-40.

[28] Pham, Duc-Hong, and Anh-Cuong Le. "Learning Multiple Layers of Knowledge Representation for Aspect Based Sentiment Analysis." Data & Knowledge Engineering (2017).

[29] Ma, Yukun, Haiyun Peng, and Erik Cambria. "Targeted Aspect-Based Sentiment Analysis via Embedding Commonsense Knowledge into an Attentive LSTM." (2018).

[30] Akhtar, Md Shad, et al. "Feature selection and ensemble construction: A two-step method for aspect based sentiment analysis." Knowledge-Based Systems 125 (2017): 116-135.

[31] Rabab'ah, Abdullateef M., et al. "Evaluating SentiStrength for Arabic Sentiment Analysis." Computer Science and Information Technology (CSIT), 2016 7th International Conference on. IEEE, 2016.

[32] Al-Ayyoub, Mahmoud, et al. "Deep learning for Arabic NLP: A survey." Journal of Computational Science (2017).

[33] Rushdi- Saleh, Mohammed, et al. "OCA: Opinion corpus for Arabic." Journal of the American Society for Information Science and Technology 62.10 (2011): 2045-2054.

[34] Abdul-Mageed, Muhammad, and Mona T. Diab. "AWATIF: A Multi-Genre Corpus for Modern Standard Arabic Subjectivity and Sentiment Analysis."LREC. 2012.

[35] Al-Kabi, Mohammed, et al. "Building a Standard Dataset for Arabic Sentiment Analysis: Identifying Potential Annotation Pitfalls" 2016 IEEE/ACS 13th International Conference of Computer Systems and Applications (AICCSA). 2016.

[36] ElSahar, Hady, and Samhaa R. El-Beltagy. "Building large arabic multi-domain resources for sentiment analysis." International Conference on Intelligent Text Processing and Computational Linguistics. Springer International Publishing, 2015.

[37] Nabil, Mahmoud, Mohamed Aly, and Amir F. Atiya. "Astd: Arabic sentiment tweets dataset." Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. 2015.

[38] Rosenthal, Sara, Noura Farra, and Preslav Nakov. "SemEval-2017 task 4: Sentiment analysis in Twitter." Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017). 2017.

[39] Al-Moslmi, Tareq, et al. "Arabic senti-lexicon: Constructing publicly available language resources for Arabic sentiment analysis." Journal of Information Science (2017): 0165551516683908.

[40] Aly, Mohamed A., and Amir F. Atiya. "LABR: A Large Scale Arabic Book Reviews Dataset." ACL (2). 2013.

[41] Nuseir, Aya, et al. "Improved Hierarchical Classifiers for Multi-Way Sentiment Analysis." International Arab Journal of Information Technology (IAJIT) 14 (2017).

[42] Al-Ayyoub, Mahmoud, et al. "Hierarchical classifiers for multi-way sentiment analysis of arabic reviews." International Journal of Advanced Computer Science and Applications (IJACSA) 7.2 (2016): 531-539.

[43] Medhaffar, Salima, et al. "Sentiment Analysis of Tunisian Dialects: Linguistic Ressources and Experiments." Proceedings of the Third Arabic Natural Language Processing Workshop. 2017.

[44] Farra, Noura, Kathleen McKeown, and Nizar Habash. "Annotating Targets of Opinions in Arabic using Crowdsourcing." ANLP Workshop 2015. 2015.

[45] El-Kilany, Ayman, Amr Azzam, and Samhaa R. El-Beltagy. "Using Deep Neural Networks for Extracting Sentiment Targets in Arabic Tweets." Intelligent Natural Language Processing: Trends and Applications. Springer, Cham, 2018. 3-15.

[46] Wahsheh, Heider, et al. "SPAR: A system to detect spam in Arabic opinions." Applied Electrical Engineering and Computing Technologies (AEECT), 2013 IEEE Jordan Conference on. IEEE, 2013.

[47] Hu, Minqing, and Bing Liu. "Mining and summarizing customer reviews." Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2004.

[48] Ganu, Gayatree, Noemie Elhadad, and Amélie Marian. "Beyond the Stars: Improving Rating Predictions using Review Text Content." Proceedings of the 12th International Workshop on the Web and Databases (WebDB). 2009.

[49] Chang, Chih-Chung, and Chih-Jen Lin. "LIBSVM: a library for support vector machines." ACM Transactions on Intelligent Systems and Technology (TIST)2.3 (2011): 27.