Framework for Affective News Analysis of Arabic News: 2014 Gaza Attacks Case Study

Huda Al-Sarhan, Majd Al-So'ud, Mohammad AL-Smadi, Mahmoud Al-Ayyoub, and Yaser Jararweh
Jordan University of Science and Technology
Irbid, Jordan
{huda.nimer, majdsoud5}@gmail.com, {maalsmadi9, maalshbool, yijaraweh}@just.edu.jo

***Abstract*—This paper aims at fostering the domain of Arabic affective news analysis through providing: (a) a benchmark annotated Arabic dataset of news for affective news analysis, (b) an aspect-based sentiment analysis (ABSA) approach for evaluating the sentimental affect of Arabic news posts on the reader, and (c) a baseline approach with a common evaluation framework to compare future research results with the baseline ones.**

Keywords—Affective News; Emotional Affect; Aspect Based Sentiment Analysis; Natural Language Processing; Arabic Dataset.

I. Introduction

With the explosive growth of social media networks, individuals and organizations are increasingly using the social media in order to improve their decision-making process. One example is concerned with the effect online contents have on the users and their reactions to them. Thus, there is a need to build automated techniques to evaluate news posts tone and sentimental affect on readers.

Researchers have started investigating the relation between the readers and the news articles in different topics. The readers are interested in detecting and tracking news stories [1], RSS feeds or news headlines [2] and classifying news [3]. Most of the news categories listed on social networks, especially Twitter, are breaking news for cities, sports, brands, accidents, disasters and political events [4, 5].

Peggy [6, 7] defined affect as a "positive and negative evaluations of an object, behavior, or idea with intensity and activity dimensions." Affective news is an interesting domain for researchers where news posts are evaluated based on their sentimental affect on readers. It is a simpler problem than the emotional affect, which includes more involved types of emotions like happiness, sadness, angry, etc. Affective news analysis is of great importance in many domains. The tone of news posts could affect readers' attitudes towards trading and stock market [8], political parties [9, 10], wars [11], etc.

News are usually written by people with experience to attract readers and followers on social networking sites. A news post can be formulated in a very emotional way by using some words and phrases that carry an affective meaning. The emotional affect of words/aspects depends significantly on text content, the context and the readers' convictions. E.g., the term 'murder' has a negative emotional affect in general, but in the context of war, it might have a positive emotional affect if it is related to the killing of an enemy, which represents a victory.

This research aims to provide an online Arabic dataset for the purposes of research related to affective news analysis. The dataset consists of Facebook posts about the 2014 Gaza attacks. The dataset is annotated while keeping in mind the sentimental affect on the reader. Instead of simply annotating each post as merely having a positive, negative or neutral sentimental affect on the reader, we perform a more detailed and more challenging annotation by looking for each aspect contained in each post and the sentimental affect the reader might have with respect to this aspect. In a previous work of ours [12], we discussed the collection and annotation of the dataset; however, at that stage, the annotation was performed by a single human annotator with very limited experience in the non-trivial task of annotation for the purposes of aspect-based sentiment analysis (ABSA) of news posts. In this paper, we revisit the same dataset and perform a more thorough annotation. Specifically, two much more experienced human annotators are employed and their separate annotation efforts are compared. To the best of our knowledge, studies on inter-annotator agreement are rare in the field of Arabic Natural Language Processing (NLP) despite their importance. Obviously, since the annotation of the dataset is modified, the baseline experiments reported in [12] have to be repeated in view of the new annotation.

The rest of the paper is organized as follows. The following section discusses the ABSA tasks under consideration. The collection and annotation of the dataset is discussed in section III. The experimental evaluation is discussed in section V. Section VI concludes this research and discusses future plans.

II. ABSA Tasks

Based on the SemEval-2014 Task 4 description [13], the following four tasks are considered.

A. Task 1: Aspect Term Extraction (T1)

Given a news post, the objective of this task is to extract all conceivable aspect terms related to Gaza domain regardless of their polarity. Annotated aspect terms examples are: (e.g. bombing/تفجير military operations/العمليات العسكرية Israeli occupation planes/طائرات الاحتلال الإسرائيلي). Note that the aspect terms should appear explicitly in the news post.

B. Task 2: Aspect Term Polarity estimation (T2)

Based on the extracted aspect terms (T1), the objective of this task is to determine the sentiment for each aspect terms and label it as neutral, positive, or negative. Instances with conflicting sentiments are not included in our study.

C. *Task 3: Aspect Category Detection (T3)*

In this task, predefined aspect categories are used to assign each news post to exactly one aspect category. Unlike aspect terms, the aspect category need not occur explicitly in the news post. They can be inferred using adjectives, sense words, or contextual meaning and not identified using aspect terms in the post.

D. *Task 4: Aspect Category Polarity Estimation (T4)*

This task is a subtask of the previous task (T3) and its objective is to determine the polarity (positive, neutral, or negative) for each identified aspect category.

III. DATASET COLLECTION AND ANNOTATION

A. *Data Collection*

The reviews that have been selected from social media are Arabic short posts containing Breaking News. They are collected from well-known Arabic news networks such as Al Jazeera and Al Arabiya. The netvizz[1] tool has been used to crawl news posts from a Facebook page called "عاجل من غزة"/"Breaking news from Gaza".

The attacks on Gaza lasted for four months and during this interval, many significant events took place. We divide this interval into eight subintervals based on the most important events that took place and divide the posts accordingly. Table I provides a description of the main events we consider.

TABLE I. BASIC EIGHT EVENTS OF THE 2014 GAZA ATTACKS

Number	Description	Start	Finish
Event #1	The actual cause of the Gaza war in 2014: the kidnapping of three young settlers on June 13. السبب الفعلي لحرب غزة 4102: عملية اختطاف الشبان المستوطنين الثلاثة يوم 01 يونيو/حزيران.	13-6-2014	30-6-2014
Event #2	The kidnapping, torturing and burning of the child Mohammed Abu Khudair. خطف وتعذيب وحرق الطفل محمد أبو خضير.	1-7-2014	6-7-2014
Event #3	Operation "Protective Edge." عملية الجرف الصامد.	7-7-2014	8-7-2014
Event #4	"Bonian Marsous" Operation. عملية البنيان المرصوص.	8-7-2014	11-7-2014
Event #5	"Al-Asf Al-Ma'kool" Battle + Failed attempts to truce. معركة العصف المأكول + محاولات فاشلة للتهدئة.	11-7-2014	20-7-2014
Event #6	The kidnapping of the Israeli soldier Aaron Shaul. خطف وأسر الجندي الصهيوني شاؤول أرون.	20-7-2014	28-7-2014
Event #7	Children massacre in Shati refugee camp + continuation of failed attempts to truce. مجزرة أطفال مخيم الشاطئ + استمرار محاولات فاشلة للتهدئة.	28-7-2014	4-8-2014
Event #8	Declare victory + Results after accepting the terms of the truce. إعلان النصر + نتائج بعد قبول شروط التهدئة.	4-8-2014	18-9-2014

Since the number of collected posts are too large to be manually annotated by us, we randomly select 20% of the news posts on each event to undergo the manual annotation process. The resulting dataset consists of 2,265 news posts annotated with aspect terms (T1), aspect term polarity (T2), aspect category (T3), and aspect category polarity (T4). The chosen posts are divided as training and testing data (80% to 20%).

B. *Annotation Process*

To annotate our dataset, the BRAT [14] tool is used, which is a web-based customizable text annotation tool. The tool properties have been adapted to provide four types of information related to the tasks at hand. The annotation has been done by two members who are experienced with annotation for ABSA. The annotation is later validated by two senior researchers in Arabic SA. All annotators are native Arabic speakers. They have received training on annotation using BRAT. They are given guidelines for the required annotation as follows.

Aspect terms and polarities. During this stage, the annotator is asked to annotate all the explicit single/multiple terms (e.g. bombing/تفجير, military operations/العمليات العسكرية, Israeli occupation planes/طائرات الاحتلال الإسرائيلي) that are related to the basic target entity (i.e., Gaza news in our case). The aspect terms are annotated as they appear in the original post even if they are misspelled. Then, the annotators are asked to give a polarity value (positive, neutral or negative) for each annotated aspect term. Fig. 1 and Fig. 2 show a sample post along with its annotation. Table II summarizes the distribution of the aspect terms over the sentiment classes.

TABLE II. ASPECT TERMS DISTRIBUTION

Domain	Positive	Negative	Neutral	Total
Aspects	4,165	4,805	685	9,655

Aspect categories and polarities. In this stage, the annotator should detect the appropriate category for each sentence and make sure that the pre-selected aspects are related to the chosen category. Four aspect categories are considered: الخطط/Plans, النتائج/Results, التهدئة/Peace and الأطراف/Parties. The most obvious category in the sentence is selected and the others are ignored. Next, each aspect category is assigned the most appropriate polarity (positive, negative, or neutral). Table III summarizes the distribution of the news posts over the considered categories and sentiment classes.

TABLE III. NEWS POSTS DISTRIBUTION

Category	Positive	Negative	Neutral	Total
Plans (الخطط)	317	453	43	**813**
Results (النتائج)	414	503	44	**961**
Peace (التهدئة)	155	110	85	**350**
Parties (الأطراف)	78	42	21	**141**
Total	**964**	**1,108**	**193**	**2,265**

The two independent annotators annotated the complete dataset before any pre-processing step. The level of agreement between the two annotators is computed. Table IV shows the conflict percentage between the two human annotations.

TABLE IV. CONFLICT BETWEEN THE TWO HUMAN ANNOTATORS

ABSA Tasks	Task 1	Task 2	Task 3	Task 4
Conflict	0.054	0.023	0.014	0.007

[1] https://apps.facebook.com/netvizz/

After studying the differences, we find that most conflicts are found in aspect terms extraction (T1), which is expected. One common source of disagreement is subsumption. E.g., in a certain news post, one annotator chooses (Martyr/الشهيد) as an aspect term, while the other annotator chooses (Child Martyr/الشهيد الطفل). As for aspect polarity estimation (T2), the differences are mainly caused by the contextual meaning of the post, not the absolute meaning of its terms. E.g., (kill/القتل prison/السجن punishment/عقوبة) are classified as negative terms according to most lexicons, while, in the case of war, it may get a positive or negative emotional affect according to the context and which side the reader is supporting. The main reason of difference in aspect category detection (T3) is due to the different understandings of the categories (Parties/الأطراف) and (Results/النتائج). Finally, the main difference in T4 is in identifying the polarity of the category (Peace/التهدئة), where, based on the context, the news posts of the Peace category do not always have a positive emotional effect and could carry negative or neutral meaning. To resolve conflicts, a third party (two expert senior researchers) reviewed both the dataset and the differences between the two annotators and served as arbitrators.

For testing the inter-annotator agreement, Cohen's Kappa measure is used as common in the literature [15]. It depends on the proportion of observations in agreement (Pr(g)) as well as the proportion in agreement equivalent to chance (Pr(e)) [15]. In our case, it measures the agreement between the two annotators, and subtracts out the agreement due to chance. Table V shows Cohen's Kappa measure for the first task (aspect terms extraction), where (a) and (d) represent the times in which the two annotators agree and (b) and (c) represent the times in which they disagree.

TABLE V. KAPPA VALUES FOR THE ASPECT TERMS EXTRACTION TASK

<table>
<tr><th colspan="2" rowspan="2">Inter-annotator agreement (Kappa measure)</th><th colspan="3">Annotator 1 Results</th></tr>
<tr><th>Aspect Term</th><th>None Aspect</th><th>Total</th></tr>
<tr><td rowspan="3">Annotator 2 Results</td><td>Aspect Term</td><td>9,364 (a)</td><td>491 (b)</td><td>9,855 (m_1)</td></tr>
<tr><td>Non-Aspect</td><td>428 (c)</td><td>28,098 (d)</td><td>28,526 (m_0)</td></tr>
<tr><td>Total</td><td>9,792 (n_1)</td><td>28,589 (n_0)</td><td>38,381 (n)</td></tr>
</table>

Fig. 1. Example of an annotated sentence in the BRAT tool.


```
<sentence id="Event4Post_0040">
    <text>عباس، ... تكلمنا مع الجانب الامريكي وطلبنا ان يوقفوا العمليات العسكرية من جانب اسرائيل
    ونحن نحاول ان نقنع حركة حماس بوقف العمليات، ولكن للاسف لم ننجح</text>
    <aspectTerms>
        <aspectTerm term="الجانب الامريكي" from="19" to="34" polarity="neutral"/>
        <aspectTerm term="العمليات العسكرية" from="52" to="69" polarity="positive"/>
        <aspectTerm term="اسرائيل" from="78" to="85" polarity="positive"/>
        <aspectTerm term="حركة حماس" from="106" to="115" polarity="negative"/>
        <aspectTerm term="وقف العمليات" from="117" to="129" polarity="negative"/>
    </aspectTerms>
    <aspectCategories>
        <aspectCategory polarity="negative" category="التهدئة"/>
    </aspectCategories>
</sentence>
```


Fig. 2. An XML snapshot that corresponds to the annotated sentence of Fig. 1.

From Table V, the proportion of observations in agreement Pr(g), the proportion in agreement due to chance Pr(e) and Cohen's kappa k are computed as follows.

$$Pr(a) = (a + d)/n \quad (1)$$

$$Pr(e) = [(n_1 /n) * (m_1 /n)] + [(n_0 /n) * (m_0 /n)] \quad (2)$$

$$k = (Pr(a) - Pr(e))/(1 - Pr(e)) \quad (3)$$

Table VI shows the Kappa values for: **1)** two human annotations, **2)** the first human annotation and the gold annotation (after being reviewed by the two senior researchers), and **3)** the second human annotation and the gold annotation. According to [16], a Kappa value of 1 shows perfect agreement, while a kappa value of 0 shows agreement due to chance. Thus, we have almost perfect agreements for the four tasks at hand.

TABLE VI. KAPPA VALUES FOR EACH PAIR OF ANNOTATIONS

	Two human annotations	Gold and first human annotation	Gold and second human annotation
Task #1	94%	95%	93%
Task #2	98%	97%	98%
Task #3	96%	95%	97%
Task #4	98%	98%	98%

We follow the dataset schema of [13, 17], where the BRAT annotation files are mapped into the compliant XML file as depicted in Fig. 2, each news post is annotated based on the three main XML tags: *text*, *aspectTerm* and *aspectCategory*. Text tag contains the original news post. AspectTerm tag contains four attributes: term (selected aspect), its polarity and the location of the aspect term in the post (from, to). AspectCategory contains two attributes: category type and the polarity of the selected category. Out of annotation phases, we have 2,265 news posts annotated and XML-formatted. They are available on request for noncommercial research.

IV. EVALUATION

The prepared dataset can be used to evaluate the sentimental affect of Arabic news posts on the reader using ABSA technique. Our dataset is tested with the baseline classifiers of [13, 17]. We also develop our own baseline.

To measure the performance of ABSA tasks, the standard performance measures are applied. For the first and third task, the F_1 measure is computed as follows.

$$F1 = \frac{2 \,.\, P \,.\, R}{P+R} \qquad (4)$$

where precision (P) and recall (R) are calculated as follows.

$$Precision\ (P) = \frac{TP}{TP+FP} \qquad (5)$$

$$Recall\ (R) = \frac{TP}{TP+FN} \qquad (6)$$

where TP is the set of true positives (relevant aspect terms for T1 or aspect categories for T3) which have been extracted from the test dataset. FP is the set of false positives (irrelevant aspect terms, aspect categories or comment categories) retrieved for the same test part. FN is the set of relevant aspects or categories not retrieved.

The simple accuracy measure is used to evaluate estimating aspect term polarity for T2 and aspect category polarity for T4. The accuracy is defined as the number of correctly retrieved polarities divided by the total number of aspect term polarity or aspect category polarity annotations.

A. SemEval Baseline

The prepared dataset is evaluated by using the baseline model. According to task 1 (T1), the results of precision, recall and F1 measures are P= 0.379, R= 0.41, and F1= 0.394, respectively. For task 2 (T2), the accuracy for aspect term polarity is 0.659. The results for task 3 (T3) are P= 0.649, R= 0.649, and F_1= 0.649. The accuracy for the category polarity estimation, which is task 4 (T4), is 0.74. these are very promising results.

B. Lexicon-Based Baseline

We propose a lexicon-based approach to address the first two tasks of ABSA. The details are as follows.

T1: Aspect term extraction:

The proposed approach is summarized in Fig. 3. It consists of the following steps: (i) manual aspect terms lexicon construction, (ii) text pre-processing, and (iii) feature extraction. For the first step, a lexicon is built by extracting all aspect terms from the training set. Then several pre-processing steps are performed including normalization, removal of news source, non-Arabic words, hyperlinks, diacritics, punctuation and special symbols, tokenization and part-of-speech (POS) tagging. Finally, in the last step, we employ the features: POS tags, N-gram, position, and pruning, enabling as described below.

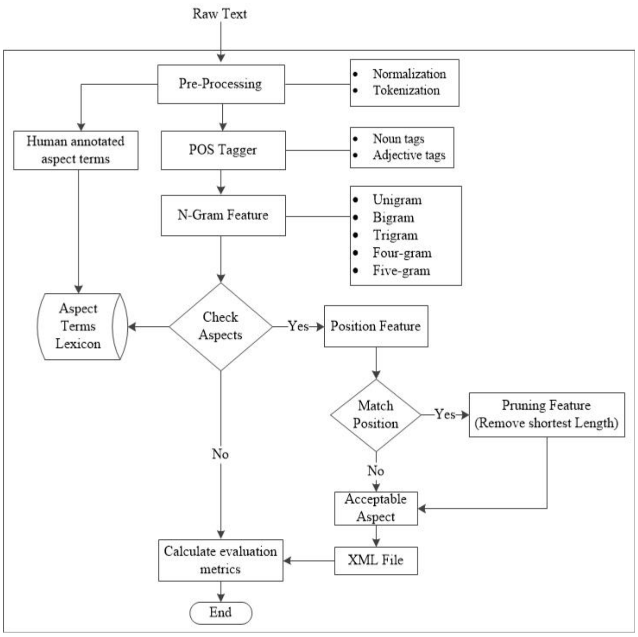

Fig. 3. Lexicon-Based Approach to extract aspect terms.

For the first feature, the POS tag for each word is computed using the AraNLP tool [18]. Note that due to the nature of the aspect terms in the prepared lexicon, the POS tags are used to retrieve words from noun/adjective categories and discard other tag types (e.g., verbs, adverb, prepositions). Table XI presents an example of a news post and how the POS tagging and the filtration step based on it alters the post.

TABLE VII. STEPS OF THE POS TAGS FILTERATION.

(1)	**Original sentence**	تكلمنا مع الجانب الأمريكي وطلبنا أن يوقفوا العمليات العسكرية من جانب إسرائيل، ونحن نحاول ان نقنع حركة حماس بوقف العمليات، ولكن للأسف لم ننجح
(2)	**Tagged sentence (complete version)**	تكلمنا/VBD مع/NN الجانب/DTNN الامريكي/DTJJ و/CC طلبنا/VBD ان/IN يوقفوا/VBP العمليات/DTNNS العسكرية/DTJJ من/IN جانب/NN اسرائيل/NNP ,/PUNC و/CC نحن/PRP نحاول/VBP ان/IN نقنع/VBP حركة/NN حماس/NNP ب/IN وقف/NN العمليات/DTNNS ,/PUNC و/CC لكن/VBP للأسف/NN ل/IN م/WP ننجح/VBP
(3)	**Tagged sentence (Keep noun / adjective tags)**	الجانب/DTNN الامريكي/DTJJ العمليات/DTNNS العسكرية/DTJJ جانب/NN اسرائيل/NNP حركة/NN حماس/NNP وقف/NN العمليات/DTNNS للأسف/NN
(4)	**Sentence after remove tags part**	الجانب الامريكي العمليات العسكرية جانب اسرائيل حركة حماس وقف العمليات للأسف

The next step after executing the POS tagging feature is applying n-gram feature to extract all the possible words/phrases of each post. An n-gram is a contiguous sequence of n words. In our algorithm, a minimum of n = 1 and a maximum of n = 5 are used. So, we create all possible sequences of 1-word, 2-words, …, 5-words phrases. Then, all these sequences are compared with the prepared aspects lexicon. If there is a matching result, the words/phrases are retrieved to check the position feature (as explained later). The example of Table VII has many n-grams, but the acceptable list with correct meanings which can be found in the prepared lexicon is: {"الجانب الأمريكي/American side", "العمليات/ operations", "العمليات العسكرية/military operations", "إسرائيل/Israel", "حماس/Hamas", "حركة حماس/Hamas", "وقف العمليات/stop the operations"}. Actually, this feature is useful since, in many cases, the human annotators select multiple contiguous words.

The successfully matched phrases are marked and their positions are checked in the original text. StartIndex and EndIndex for each returned word/phrase are computed. It is an important step in the algorithm because the lexicon contains some aspect terms that might be parts of other phrases, e.g., {"operations/العمليات", "military operations/العمليات العسكرية"}, {"الطائرات/Planes", "الطائرات الحربية/Warplanes", "الطائرات الحربية الاسرائيلية/Israeli warplanes"}, {"حماس/Hamas", "حركة حماس/Hamas"}. In this example, three groups contain the retrieved aspect terms from the different posts and can be found in the prepared aspects lexicon. The first two groups have the same StartIndex and the last group has the same EndIndex. The essential question is which one of the retrieved aspects in each group is the correct aspect? Inspecting the StartIndex and EndIndex (position) of such words/phrases can help in answering this question accurately.

Finally, for the pruning step, the phrases with the same StartIndex or the same EndIndex (same position) are captured. Then, all aspects with shorter lengths are neglected. In contrast, the aspect that has the longest length are stored in the final XML file. For the example of the previous paragraph, the correctly returned aspects from each group are {"military operations/العمليات العسكرية"}, {"الطائرات الحربية الاسرائيلية/Israeli warplanes"}, {"حركة حماس/Hamas"}.

T2: Aspect term polarity estimation:

A lexicon-based approach is proposed for T2 as follows:

1. A simple sentiment lexicon file is built from the training set. This lexicon is limited and is just related to the 2014 Gaza attacks posts. Our sentiment lexicon consists of six different values where each value has a specific meaning as follows:

a. In all training dataset, the aspect terms that belong exclusively to one sentiment class (positive, negative or neutral) are given values (1, -1, 0) respectively. This case is considered as the highest priority. Examples include {"الأبطال/Heroes: positive"} → (1), {"مجازر/Massacres: negative"} → (-1), {"أوضاع/Situations: neutral"} → (0).
b. If an aspect term belongs to more than one sentiment class in the training data, the frequency of its appearance in each class is computed. The class that has the highest frequency is the class assigned to this aspect. If the highest frequency is for the positive class, then the label value is (11). While if the highest frequency is for negative class, then the aspect provides value (-11). Finally, if the highest frequency is for neutral class, then the aspect provides value (00)). This case is considered as the second highest priority. E.g., {"الطاقم الطبي/Medical staff: positive frequency"} → (11), {"أشلاء/Remains: negative frequency"} → (-11), {"اجتماع/Meeting: neutral frequency"} → (00).
c. If the frequency is similar for all classes, the given aspect term is removed from the sentiment lexicon because we cannot give an accurate decision about its polarity. E.g., "المبادرة المصرية/Egyptian initiative" → Frequency value for each class (positive, negative, neutral) = (1, 1, 1).

2. An external open source sentimental lexicon (Positive & Negative Arabic sense words) is used.[2] In addition, some general military terms that are collected from the public news articles have been added such as حرب/war, أسر/capture, اغتيال/assassination, قتلى/killed, جرحى/wounded, أسرى/captured, etc. The sense words are labeled by (+1) for the positive words and (-1) for the negative words. In this way, the external and general sentiment lexicon is ready to be integrated into our approach.

After finishing the preparation of the required lexicons, they are considered based on certain priority rules as follows:

1. Each post consisting of noun and noun phrases is retrieved and split it into tokens. In addition, the human annotated aspect terms are read.
2. The sentiment lexicon that was created from the training dataset is opened. Then, each aspect term that belongs to a specific post is compared with terms in the lexicon. If found, then we return its polarity value, which is either (1, -1, 0) → (positive, negative, neutral) for the highest priority, or (11, -11, 00) → (positive frequency, negative frequency, neutral frequency) for the second highest priority.
3. If the aspect term does not exist in the sentiment lexicon, then we check the words (tokens) around each aspect term. In other words, from the given aspect term, we move one word/token backward and one word/token forward and check the tokens' polarities. If they are found in the sentiment lexicon, we return the token values and the cases we have are (1, -1, 0) or (11, -11, 00) → (positive Before/After, negative Before/After, neutral Before/After). This case is referred to as the third highest priority.
4. The external lexicon that contains positive and negative words is opened to check if the given aspect term is found or not. If yes, then the aspect is assigned with different label like (1, -1) → (neg, pos) to distinguish the priority. This case has the fourth highest priority.
5. The last check also involves moving one word/token backward and one word/token forward around each aspect in an attempt to find a polarity for the adjacent tokens. This case has the least priority which is the fifth

[2] https://goo.gl/m4JVKF

and it is labeled as (0, 1) → (pos Before/After, neg Before/After).

6. In some cases, aspect terms have a blank polarity value because all previous cases did not find a match, either aspect term does not exist in two lexicons or the token before and after aspect term as well cannot be found in both lexicons. In this case, the algorithm of repairing empty polarities is applied and it has three cases.
   a. Fill the empty polarity value of the current aspect term with the polarity of the previous aspect term in the same post.
   b. If there is no previous value, then fill the empty polarity value with the polarity of the next (after) aspect term in the same post.
   c. The worst case is when all aspects in some posts are still empty, despite all of our attempts. In this case, all the empty polarities will be assigned with the sentiment class that has the highest frequency in the whole training file.

The results of the lexicon-based approaches are as follows. For T1, the results are 0.485, 0.546 and 0.514 for the precision, recall and F_1, respectively, which represent improvements of 11%, 14% and 12%. Although the lexicon-based and baseline methods rely on using simple lexicons to retrieve the terms from the training part of the database, the improvement ratio for the lexicon-based method is due to applying extra features that are not implemented in the baseline model. As for T2, the resulting accuracy is 0.673 which is only about 1% better than the baseline model.

V. Conclusion and Future Work

The goal of this research is to evaluate news affect on reader by providing a dataset of Arabic text in news domain. News posts and their comments have been collected during Israel-Gaza conflict of 2014 from well-known Arabic news networks such as Al-Jazeera and Al-Arabiya. The dataset consists of 2,265 news posts. The dataset has been prepared and annotated to provide important information for all tasks related to the ABSA based on SemEval-2014 Task 4 [13] guidelines. The four tasks under consideration are: aspect terms extraction (T1), aspect term polarity identification (T2), aspect category selection (T3) and aspect category polarity identification (T4). Then, the dataset (represented in XML format) is evaluated by executing [13]'s baseline model for four ABSA tasks. The common accuracy measures: precision, recall, F_1 and accuracy, are used. Therefore, any proposed approach concerned with studying the affective news by using ABSA technique can be compared with the results of the baseline model. Finally, we proposed enhanced lexicon-based approaches for aspect term extraction (with 12% improvement in the F_1 measure) and aspect term polarity estimation (with 1% improvement in accuracy).

As future work, we are working on extending the proposed lexicon-based approach to the other tasks of ABSA. We also plan to extend our dataset of news posts to include the comments made by the readers of these posts. The goal is to measure the level of agreement in the sentiment value between news posts affect and users' comments (replies).

Acknowledgment

This research is partially funded by Jordan University of Science and Technology, Research Grant Number: 20150164.

References

[1] J. Allan, J. Carbonell, G. Doddington, J. Yamron, and Y. Yang, "Topic detection and tracking pilot study final report," In the DARPA Broadcast News Transcription and Understanding Workshop, 1998.

[2] X. Wan and J. Yang, "Collabsum: exploiting multiple document clustering for collaborative single document summarizations," Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR '07, pp. 143–150, 2007.

[3] H. Kwak, C. Lee, H. Park, and S. Moon, "What is Twitter, a social network or a news media?," Proceedings of the 19th international conference on World wide web - WWW '10, pp. 591–600, 2010.

[4] F. Sebastiani, "Machine learning in automated text categorization," CSUR ACM Comput. Surv. ACM Computing Surveys, vol. 34, no. 1, pp. 1–47, 2002.

[5] P. Farhi, "The twitter explosion," American Journalism Review, vol. 31, no. 3, pp. 26–31, 2009.

[6] M. D. Munezero, C. S. Montero, E. Sutinen, and J. Pajunen, "Are They Different? Affect, Feeling, Emotion, Sentiment, and Opinion Detection in Text," IEEE Trans. Affective Comput. IEEE Transactions on Affective Computing, pp. 101–111, 2014

[7] P. Thoits, "The sociology of emotions," Annual review of sociology, pp. 317–342, 1989.

[8] M. Gentzkow and J. Shapiro, "What Drives Media Slant? Evidence from U.S. Daily Newspapers," no. w12707, pp. 35–71, 2006.

[9] M. Eshbaugh-Soha, "The Tone of Local Presidential News Coverage," Political Communication, vol. 27, no. 2, pp. 121–140, 2010.

[10] S. Farnsworth and R. Lichter, "The nightly news nightmare: Media coverage of US presidential elections, 1988-2008," Rowman & Little field Publishers, 2010.

[11] J. Cho, M. P. Boyle, H. Keum, M. D. Shevy, D. M. Mcleod, D. V. Shah, and Z. Pan, "Media, Terrorism, and Emotionality: Emotional Differences in Media Content and Public Reactions to the September 11th Terrorist Attacks," Journal of Broadcasting & Electronic Media, vol. 47, no. 3, pp. 309–327, 2003.

[12] M. Al-Smadi, H. Al-Sarhan, M. Al-Ayyoub, Y. Jararweh and E. Benkhelifa, "Using Aspect-Based Sentiment Analysis to Evaluate Arabic News Affect on Readers," Proceedings of the 8th IEEE/ACM International Conference on Utility and Cloud Computing (UCC), 2015.

[13] M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos, and S. Manandhar. "SemEval-2014 Task 4: Aspect Based Sentiment Analysis". In Proceedings of the 8th International Workshop on Semantic Evaluation, SemEval 2014, Dublin, Ireland.

[14] P. Stenetorp, S. Pyysalo, G. Topić, T. Ohta, S. Ananiadou, and J. Tsujii. "BRAT: a web-based tool for NLP-assisted text annotation". In Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics (EACL '12), pp. 102-107, 2012.

[15] A. J. Viera, and J. M. Garrett, "Understanding interobserver agreement: the kappa statistic," Fam Med, vol. 37, no. 5, pp. 360–363.

[16] D. G. Altman, "Practical Statistics for Medical Research," Chapman and Hall, 1991.

[17] M. Al-Smadi, O. Qawasmeh, B. Talafha, and M. Quwaider, "Human Annotated Arabic Dataset of Book Reviews for Aspect Based Sentiment Analysis 2015," Future Internet of Things and Cloud (FiCloud), 2015 3rd International Conference on, pp. 726–730, 2015.

[18] M. Althobaiti, U. Kruschwitz, and M. Poesio, "AraNLP: A java-based library for the processing of Arabic text." In the 9th Language Resources and Evaluation Conference (LREC 2014), 2014.