Open Issues in the Sentiment Analysis of Arabic Social Media: A Case Study

Samhaa R. El-Beltagy
Center for Informatics Science
Nile University
Giza, Egypt
selbeltagy@nileuniversity.edu.eg

Ahmed Ali
Faculty of Information and Computer
Cairo University,
Giza, Egypt
ahmedrali2@gmail.com

Abstract— **With the rapid increase in the volume of Arabic opinionated posts on different microblogging mediums comes an increasing demand for Arabic sentiment analysis tools. Yet, research in the area of Arabic sentiment analysis is progressing at a very slow pace compared to that being carried out in English and other languages. This paper highlights the major problems and open research issues that face sentiment analysis of Arabic social media. The paper also presents a case study the goal of which is to investigate the possibility of determining the semantic orientation of Arabic Egyptian tweets and comments given limited Arabic resources. One of the outcomes of the presented study, is an Egyptian dialect sentiment lexicon**

Keywords—Opinion Mining; Sentiment Analysis; Arabic; Social Media; Sentiment Lexixons; Microblogs

I. INTRODUCTION

Sentiment analysis has been gaining lot of attention over the past number of years due to the important role that it can play in many different areas among which are marketing and politics. The wide spread use of microblogging services has also led to a wide spread availability of opinionated posts, making research in the area more viable and important. While there is no shortage in literature of works addressing English sentiment analysis, there is only a handful of published works addressing Arabic sentiment analysis; this, despite the fact that Arabic speaking nationals are increasingly making use of microblogging services. A study prepared and published by Semiocast, which is a French based company, has revealed that Arabic was the fastest growing language on Twitter in 2011, and was the 6th most used language on Twitter in 2012[1]. The same study revealed that the number of twitter users in Saudi Arabia almost doubled (grew by 93%) in the span of 6 months in 2012 and that Riyadah is now the 10th most active city on Twitter. A breakdown of Facebook(FB) users by country, places Egypt with 11,804,060 users as the Arabic speaking country with the largest number of FB users, and ranks it at 20 among all countries of the world. Saudi Arabia follows with a user base of 5,760,900 and an overall rank of 33[2]. Another interesting study has shown that Arabic speaking countries, specifically, Lebanon, Tunisia, Egypt and Jordan, have the most politically vocal social networks[3]. While these figures serve to highlight the importance of carrying out Arabic sentiment analysis, the field is progressing at a very slow pace. The goal of this paper is to examine and highlight the main challenges facing Arabic sentiment analysis of social media.

The rest of this paper is organized as follows: Section II presents the background for this work, section III, discusses Arabic specific challenges when carrying out sentiment analysis of social media, section IV presents the case study, section V, presents experiments carried out to evaluate the approach taken in the case study and their results, and finally section VI concludes this paper.

II. BACKGROUND

Sentiment analysis or opinion mining can be defined as the task of determining the semantic orientation of an opinion holder ***H*** on an object or a feature of an Object ***O***. The semantic orientation(SO) can be either positive, negative or neutral. The problem has many facets, with some work focusing only on determining the SO of a passage of text, and other work addressing the problem of extracting the opinion holder or the object feature being described. When only considering the task of SO determination, two main approaches have been reported in literature. The first approach utilizes a sentiment lexicon as well as a part of speech (POS) tagger to determine whether the orientation of some given text is positive, negative or objective. The sentiment lexicon usually contains a list of opinionated words and their strength as either negative, positive and objective terms. The second approach relies on the availability of text tagged with SO which is then used to train a classifier to determine the SO or new or unseen text within the same domain as the tagged corpus. The main problem with the machine learning approach is that it degrades rapidly if applied on any other domain on which it was not trained [1].

Work addressing sentiment analysis in Arabic is very limited. One of the earliest works on Arabic opinion mining is that presented in [2] in which the goal was to mine Arabic business reviews. In this work, the authors used a seed set of 1600 words (600 positive, 900 negative, and 100 neutral). The seed words were used within an Arabic similarity graph built using a large web corpus to determine the polarity of neighboring terms using a labeling propagation mechanism. The output of this process was a lexicon that was one of the main components of the developed sentiment analyzer. More recently Abdul-Mageed et al [3] presented a system for carrying out sentence level subjectivity and sentiment analysis for modern standard Arabic (MSA) text. Abdul-Mageed and Diab also presented a system for learning an Arabic lexicon[4], but again, the focus was on MSA terms.

Work that targets Arabic social media, is even more limited. SAMAR [5], tackles the problem of sentiment analysis in social media from a mostly linguistic perspective. The system is based on support vector machine (SVM) classifiers and carries out SO determination in two steps. In the first step, a classifier is used to distinguish between subjective (opinionated) cases from objective ones. In the second step, another classifier is used to determine whether the subjective input is positive or negative. Neutral and mixed cases are not addressed by the system. Some of the features used by the classifiers include morphological features, part of speech (POS) tags, and matches made with entries in an adjective polarity lexicon which simply classifies adjectives as either positive or negative. The dialectical performance of the system was evaluated using the Tagreed dataset which consists of 3015 Arabic divided into 1466 written in MSA 1549 tweets written in different dialects. 80% of each of the datasets were used for training, 10% for developments, and 10% for testing. The highest accuracy reported through the dialect-specific sentiment experiments was 71.15% with an F-score of 29.4% for positive cases and an F-score of 81.8% for negative ones. The fact that the dialect specific dataset consists mostly of negative tweets, has balanced out the low positive F-score when assessing the overall accuracy. The work of [6] examines the effect of preprocessing on the task of determining the semantic orientation of Tweets written in the Egyptian dialect. To do so, the authors makes use of a corpus that they built and annotated and which consists of 500 positive tweets and 500 negative tweets. 600 of the annotated tweets (300 positive and 300 negative) are used to build a lexicon. Weights for the terms in the lexicon were assigned based on the frequency of their occurrence in a positive context relative to their occurrence in a negative context within the used 600 tweets. Using a SVM classifier on the remaining 400 tweets and after experimenting with several pre-processing steps, the best accuracy obtained was 78.8%.

III. SENTIMENT ANALYSIS OF SOCIAL MEDIA: ARABIC SPECIFIC CHALLENGES

The main difficulty of performing sentiment analysis in Arabic social media lies in the fact that communication within the social media context is carried out using "spoken" or colloquial Arabic rather than the more formal Modern Standard Arabic (MSA). Not only is the vocabulary of colloquial Arabic different than that of MSA, the structure of the sentences is much more random which is why parsing this text poses a major challenge. The following subsections examine some of the problems that are bound to face anyone working on colloquial Arabic sentiment analysis.

A. Unavailability of colloquial Arabic parsers

Arabic is a morphologically complex language that has a high inflectional and derivational nature. Parsing modern standard Arabic (MSA) is an already complex task towards which many efforts have been directed. However, colloquial Arabic differs from MSA phonologically, morphologically, and lexically and does not have a standard orthography complicating the task of building morphological analyzers and part of speech taggers for it even further. Many sentiment analysis tools require the use of these tools to accomplish their tasks. While some work has been carried out to address this problem, examples of which are presented in [7] and [8], the accuracy of available tools is still not comparable to accuracies obtained for MSA and other foreign languages.

B. Unavailability of Sentiment Lexicons

Sentiment lexicons containing opinion terms, along with their polarity and strength are an essential part of any sentiment analysis tool. There are currently no publically available colloquial Arabic sentiment lexicons of which we know of, prompting the need to build these.

A lot of work has been carried out to address the question of how to build a polarity lexicon for many languages including but not limited to German, Dutch, Spanish, Chinese, and Japanese, but most notably, English. The work described in [12] for example, has proposed an unsupervised algorithm for inferring the polarity of phrases that have adjectives and adverbs within a POS-tagged corpus. Polarity is calculated based on Pointwise Mutual Information between unknown phrases and the words "excellent and poor". The approach suffers from several drawbacks if considered for application within an Arabic microblogging context: 1. it relies on the existence of a huge POS-tagged corpus, which in the case of Arabic microblogs is particularly challenging as the language used within these blogs is highly unstructured and a POS tagger that can actually work on these with an acceptable degree of accuracy, is yet to be developed. 2. The approach only targets adjectives and adverbs even though verbs and nouns can also play a role in determining polarity.

The work presented in [7] proposes an approach for building a large scale Arabic sentiment lexicon. To do so, the authors expand on a Modern standard Arabic (MSA) polarity lexicon of 3225 adjectives which they built manually. Extension is achieved by using a number of existing English lexicons including SentiWordNet[13], which is the most famous and most widely used English polarity lexicon. The authors reported having problems with both coverage and the quality of some of the entries. They also stated that they have not tested the system for the task of sentiment analysis. But even without the reported problems, this approach will still be incapable of encompassing slang, dialectical Arabic, and multiword expressions. It also focuses only on adjectives and does not seem to be concerned with assigning polarity scores to different terms. So building a colloquial Arabic polarity lexicon remains an open research area.

C. The need for person name recognition

In most languages, carrying out named entity recognition (NER) is not a requirement for determining the semantic polarity of some given text. NER becomes an important part of sentiment analysis only when the task is to identify an opinion holder and the object of the opinion. However, in the context of Arabic social media, where person names are very commonly used, and specially in the absence of POS taggers, person name recognition becomes a requirement even for the task of determining semantic orientation. The reason for this is that the majority of Arabic first names, and to a lesser extent family names, are derived from Arabic adjectives that can be easily confused for sentiments. Some Arabic male names that demonstrate this point include [Adel, Nabil, Said, and Hakim] the translations of which are [Just, Noble, Happy, and Wise]. Example of female names include [Gamila, Latifa, Sara, and Wafia] which translate to [Beautiful, Nice, Happy, and Loyal].

D. Handling compound phrases and idioms

Compound phrases and idioms are very commonly used by Arabic speakers in social media settings to express their opinions. While some of those are common amongst all Arabic speakers, many vary from one country to another. When analyzing one hundred opinionated comments taken from the famous Egyptian "كلنا خالد سعيد" page (a page that is said to have played an important role in Egypt's Jan. 25th revolution and which is known for its continuing active political role) thirty six of those comments (36%) contained at least one idiom or compound phrase. The comments were collected from seven different posts, and were taken in the order that they appeared in order to eliminate any bias. Further manual analysis revealed that the sentiment reflected by 24 of those expressions cannot be determined by any of their constituent words. An example of an idiom that is negative and whose constituent words convey a negative sentiment is "يقتل القتيل و يمشي في جنازته" the literal translation of which is "he kills the victim and walks in his funeral". The words: kill, victim, and funeral, all point towards a negative sentiment. An example of a negative expression whose constituent words do not convey a negative sentiment is "يا واد يا مؤمن" the closest translation of which is "oh, you believer you" and which is actually an expression that belittles someone pretending to be pious. Other examples include "لا ياشيخ" which according to Google translate corresponds to "O Sheikh", but which in the Egyptian Arabic context reflects an extreme disbelief in what someone is saying, and hence a negative sentiment even though none of its constituent words are themselves negative. Similarly, "حسبي الله ونعم الوكيل" which Google was unable to translate, and which roughly means, "I can only depend on God", usually reflects a very negative sentiment as it carries an implicit prayer for God to take revenge on those who have carried out an injustice. This is also an example of an expression that is shared among all Arabic speakers. Again, none of the Arabic or English words that make up this phrase are negative. The phrase "بيلحس جزم" or "Shoes licker" is used to describe someone as a hypocrite. The same is also true for positive sentiments. For example, the phrase "زي العسل" which translates to "like honey", is often used to say that someone is very sweet.

The actual usage of idioms and expressions varies from one forum to another and from one culture to another and depends heavily on the subject being discussed, but there is no question that in general Arabic speakers employ those in social media settings more than English speakers.

IV. Case Study: Esrabilishing the Semantic Orientation of Egyptian Tweets

The aim of the presented case study was to carry out a preliminary investigation to determine whether the task of establishing semantic orientation of Egyptian Arabic microblogs in light of the outlined challenges is doable. The adopted approach is comprised of three steps: first, building a sentiment lexicon, second, assigning weights to words in the lexicon and third, determining semantic orientation based on the sentiment lexicon. As stated in section 3.C, person name recognition is a very important step when carrying out Arabic sentiment analysis. For this purpose, a person named entity recognizer that does not require any Arabic parsers was developed and used. Testing this system on the ANERcorp benchmark dataset, resulted in an F-score of 81% [9]. The following subsections, describe in details the first two steps. Several methods were followed for the third step. These are described in the next section (Experiments and results).

A. Building the sentiment lexicon

The goal of building a sentiment lexicon was to have a tangible component to use for the purpose of determining semantic orientation. Towards this end, a seed sentiment lexicon comprised of 380 words was manually constructed and used to collect more sentiment terms. Each of the terms in the lexicon was tagged as being either a verb, adjective, noun, adverb, or idiom/compound. The expansion of the seed lexicon was based on the idea that sentiment terms often appear with other sentiment terms that have the same polarity. A simple way for detecting these is looking for conjugated patterns. For example, assuming that the positive term "محترم" or "respectable" exists in the seed lexicon, if the phrase "محترم و مؤدب" or "polite and respectable" is encountered, it will be safe to assume that "مؤدب" is a sentiment term that also has a positive polarity. Users of twitter sometimes use it to direct compliments or insults to other twitter users. So anther way for collecting sentiment terms is making use of this observation. Using the same terms as above and encountering the pattern "يا مؤدب يا محترم" which is not easily translatable to English because this form of address is particular to Arabic but which roughly translates to "you polite you respectable one", it is also possible to deduce that "مؤدب" is a sentiment term that also has a positive polarity. Based on these observations, the algorithm used to expand the seed lexicon is shown in Fig. 1. After collecting all possible candidates and their counts using the outlined procedure, these were filtered manually to exclude any incorrectly learned candidate terms. Excluded terms were kept in a falseCandidates list and the procedure was repeated again using the newly acquired lexicon terms. Candidate terms that matched with terms in the falseCandidates were excluded automatically. It is important to note that only single term sentiments rather than compound phrases are learned in this manner. After applying this procedure a number of times, a sentiment lexicon of 4,392 terms was constructed. This lexicon is available for download from: http://tmrg.nileu.edu.eg/.

B. Assigning weights to lexicon terms

Preliminary experiments using just the polarity of terms in the lexicon to determine the SO orientation of a short piece of text, revealed the need to assign weights to lexicon terms. Intuitively, the strength of opinionated words usually differs from one term to another. In order to assign weights to lexicon terms, two weight assignment algorithms were devised. The effect of using each, is presented in the next section. Each of the two devised algorithms requires a big collection of text to derive weights from. To build this collection, up to 1000 tweets per lexicon entry were collected using Twitter's search API. The total number of tweets that were collected as a result was 2,700,000 tweets. The pseudo code for each of the followed methods is shown in figures Fig.2 and Fig. 3 respectively.

```
for each term w ∈ lexicon
  use twitter api to search for term (w)
  for each tweet having term w
    if pattern (w و c) or (c و w) or (يا w يا c) or (يا
        c يا w)  appears in the body of the tweet
        consider c as a candidate sentiment term
        with the same polarity as w
```

Fig. 1. Pseudo code for expanding lexicon entries

```
For each entry e ∈ lexicon
    Search collected tweets for those that contain e and store in
matches
    create an oppositePolarityCounter and initialize it to 0
    create a validTexts counter and initialize it to 0
    For each tweet  t ∈ matches
        detect +ve and -ve terms in t
        If t  has no opinion terms other than e,  discard it
        Else
            increase validTexts by 1
            If the text has opinion terms of opposite polarity
                increase oppositePolarityCounter by 1
        end for
  set  weight(e)=(validTexts - oppositePolarityCounter)/
                        validTexts //sign is determine by polarity of e
```

Fig. 2. Pseudo Code for first weight assignment algorithm

```
For each entry e ∈ lexicon
If e is positive set weight(e) = 1
   Else set weight(e) = -1
Search collected tweets for those that contain e and store in
matches
create a validTexts counter and initialize it to 0
 For each tweet  t ∈ matches
        If e is negated in t, discard t
        Else
          calculate polarity of t by summing weights of all opinion
          words in t
           if t.length >=3 & t .polarity <Ω  & t.polarity > -Ω, discard
           t    //Ω was set to 3
           Else increase validTexts by 1
End for
If e is negative
    n_count =  get countOfNegativeTweets // Tweets with polarity <0
    set weight(e) = - (n_count / validTexts)
 If e is positive
     p_count =  get countOfPositiveTexts  // Tweets with polarity >0
     set weight(e) =  (p_count / validTexts)
End for
Calculate negativeAvgweight and positiveAvgWeight
Assign (negativeAvgweight * β) to  -ve idiom terms
Assign (positiveAvgWeight * β) to  +ve idiom terms
Assign negativeAvgweight to  -ve terms for which a weight was not
    calculated
Assign positiveAvgWeight to  +ve terms for which a weight was not
    calculated
```

Fig. 3. Pseudo Code for second weight assignment algorithm

β in Fig 3. is a boosting factor assigned to compound expressions and idioms; it was set to 1.75. Examples of variations in weight assignment from one algorithm to another are shown in table 1.

TABLE I. EXAMPLES OF WEIGHTS CALCULATED USING BOTH METHODS

Terms	Polarity	English equivalent	***w*** using 1st Algorithm	***w*** using 2nd Algorithm
بحبك	positive	I love you	0.39	0.94
حب	positive	love	0.14	0.82
محترم	positive	respectable	0.12	0.86
يهدد	negative	Threatens	-0.92	-0.92
تحبس	negative	Locks up	-0.78	-0.95
اعتقل	negative	Arrested	-0.61	-1

V. EXPERIMENTS AND RESULTS

A. *Used Data Sets*

Two data sets were used for experimentation purposes. The first dataset (Twitter dataset) was built by collecting 500 random tweets and manually annotating them with their semantic orientation. The result of this process was that 310 tweets were classified as negative, 155 as positive 35 as neutral. Similarly, the second dataset (Dostour dataset) was built by randomly collecting 100 comments made on articles of the newly proposed Egyptian constitution and annotating those with their semantic orientation. 38 comments were classified as negative, 40 as positive and 22 as neutral. Both dataset are available for download form http://tmrg.nileu.edu.eg/

B. *Automatically calculating semantic orientation*

We have experimented with two different methods for deriving the semantic orientation of a piece of text. Both methods rely on indentifying sentiment words in the piece of text. Also, in both methods, person names are identified and excluded from any lexicon matching attempts. Negating a term reverses its polarity and an intensifier for a term results in an amplification of its weight by multiplying it by a factor ***f***. Positive terms always have a positive weight and negative terms have a negative weight. The first method (Sum) is a straightforward one, in which the polarity of a given text is calculated by adding up the weights of negative and positive terms, and the polarity is determined by the sign of the obtained number. The pseudo code for this method is presented in Fig. 4.

TABLE 2. RESULTS USING A UNIFORM WEIGHT OF 1 FOR +VE TERMS AND –1 FOR –VE TERMS

	-ve			+ve			neutral			Accuracy
	P	R	F	P	R	F	P	R	F	
Twitter Data set (500 tweets)										
Sum (Ω = 0, 0.3, 0.2, 0.1)	**95.1**	**81.3**	**87.7**	**88.4**	**69.0**	**77.5**	**21.1**	**68.6**	**32.2**	**76.6**
Sum (Ω= avg)	98.6	47.1	63.8	91.9	21.9	35.4	10.2	91.4	18.3	42.4
DP	**95.1**	**81.3**	**87.7**	**88.4**	**69.0**	**77.5**	**21.1**	**68.6**	**32.2**	**76.6**
Dostour Data set (100 comments)										
Sum (Ω = 0, 0.3, 0.2, 0.1)	86.7	**34.2**	**49.1**	81.8	**67.5**	**74.0**	**34.6**	**81.8**	**48.7**	**58.0**
Sum (Ω= avg)	**100**	10.5	19.1	**100**	35	51.9	26.8	100	42.3	40.0
DP	86.7	**34.2**	**49.1**	81.8	**67.5**	**74.0**	**34.6**	**81.8**	**48.7**	**58.0**

TABLE 3. RESULTS BASED ON WEIGHTS CALCULATED USING THE 1ST ALGORITHM

	-ve			+ve			neutral			Accuracy
	P	R	F	P	R	F	P	R	F	
Twitter Data set (500 tweets)										
Sum (Ω = 0)	86.2	**90.7**	88.4	**87.7**	**69.7**	**77.4**	**40.0**	57.1	**47.1**	**81.8**
Sum (Ω = 0.3)	89.2	88.1	88.6	84.0	40.7	54.8	17.6	60.0	27.3	71.4
Sum (Ω = 0.2)	88.0	90.0	**89.0**	84.2	44.5	58.2	19.8	57.1	29.4	73.6
Sum (Ω = 0.1)	86.9	90.0	88.4	69.7	**69.7**	69.7	36.4	57.1	44.4	81.4
Sum (Ω= avg)	**93.1**	73.6	82.2	84.2	44.5	58.2	13.3	65.7	22.1	64.0
DP	67.1	94.8	78.6	80.0	7.7	14.1	42.6	57.1	48.7	76.6
Dostour Data set (100 comments)										
Sum (Ω = 0)	**71.4**	**65.8**	**68.5**	**68.4**	**65.0**	**66.7**	**44.4**	**54.6**	**49.0**	**63.0**
Sum (Ω = 0.3)	68.8	57.9	62.9	65.2	37.5	47.6	35.6	72.7	47.8	53.0
Sum (Ω = 0.2)	68.8	57.9	62.9	71.4	50.0	58.8	40.0	72.7	51.6	58.0
Sum (Ω = 0.1)	70.6	63.2	66.7	65.0	65.0	65.0	42.9	54.6	48.0	62.0
Sum (Ω= avg)	81.3	34.2	48.2	71.4	50.0	58.8	33.9	86.4	48.7	52.0
DP	41.8	73.7	53.3	50.0	7.5	13.0	44.4	54.6	48.9	43.0

```
SO_w = 0;          //Semantic orientation weight
   for each sentiment_term ti ∈ text
         SO _w = SO_w + weightOf(ti)
   if SO_w > Ω Semantic_Orientation = "positive"
   else if SO_w < Ω Semantic_Orientation = "negative"
else   Semantic_Orientation = "neutral
```

Fig. 1. The Sum method for calculating semantic orientation

The second method (double polarity or DP) considers each term in the lexicon as having both a positive and a negative weight. So for example, if a positive term in the lexicon has a weight of 0.8, its negative weight will be -1 +0.8 which is -0.2. Similarly, a negative term with a weight of -0.6 will have a positive weight of 0.4. In this method, a positive score is obtained by summing all positive weights in the input text, and a separate negative score is calculated by summing all negative weights in the input text. Whichever absolute weight is greater, determines the polarity of the text. The pseudo code for this method is presented in Fig. 5.

```
positive_Score = 0;
negative_score = 0;
   for each sentiment_term ti ∈ text
        positive_Score = positive_Score + positive_weightOf(ti)
        negative_score = negative_score + negative_weightOf(ti)
end for
if positive_Score > | negative_score| then Semantic_Orientation =
    "positive"
else if positive_Score == | negative_score| then
    Semantic_Orientation = "neutral"
 else   Semantic_Orientation =  "negative"
```

Fig. 2. The double polarity (DP) method for calculating semantic orientation

Tables 2 through 4, show the results of using both methods with different Ω parameters for the Sum method. Each of the tables shows the results based on one of three different lexicon weight configurations. Ω= avg indicates that Ω is assigned the average weight calculated based on entries in the lexicon. Table 2, shows the results based on using a constant weight of 1 for positive terms and –1 for negative terms. Table 3 shows the results based on calculating weights using the 1st algorithm described in the previous section while Table 4 present the results based on calculating weights using the 2nd algorithm.

As can be seen from the tables, using a weight calculation method offers an improvement over the simple assignment of a uniform weight to positive and negative lexicon entries. The

TABLE 4. RESULTS BASED ON WEIGHTS CALCULATED USING THE 2ND ALGORITHM

	-ve			+ve			neutral			Accuracy
	P	**R**	**F**	**P**	**R**	**F**	**P**	**R**	**F**	
Twitter Data set (500 tweets)										
Sum (Ω = 0)	90.8	86.1	88.4	80.7	80.7	80.7	39.2	57.1	46.5	82.4
Sum (Ω = 0.3)	93.6	84.2	88.6	86.0	75.5	80.4	23.5	57.1	33.3	79.6
Sum (Ω = 0.2)	93.5	84.5	88.8	86.0	75.5	80.4	23.8	57.1	33.6	79.8
Sum (Ω = 0.1)	93.2	84.5	88.6	86.0	75.5	80.4	24.1	57.1	33.9	79.8
Sum (Ω= avg)	94.0	70.7	80.7	86.6	66.5	75.2	15.5	65.7	25.1	69.0
DP	**89.5**	**90.3**	**89.9**	**85.6**	**76.8**	**81.0**	**41.7**	**57.1**	**48.2**	**83.8**
Dostour Data set (100 comments)										
Sum (Ω = 0)	76.0	50.0	60.3	62.5	75.0	68.2	44.4	54.6	49.0	61.0
Sum (Ω = 0.3, 0.2, 0.1)	75.0	47.4	58.1	67.5	67.5	67.5	33.3	54.6	41.4	57.0
Sum (Ω= avg)	77.8	36.8	50.0	67.6	62.5	64.9	28.9	59.1	38.8	52..0
DP	**72.7**	**63.2**	**67.6**	**67.5**	**67.5**	**67.5**	**44.4**	**54.5**	**49.0**	**63.0**

second weight calculation method using the double polarity approach for determining semantic orientation seems to yield the best results with 83.8% accuracy on the Twitter dataset and 63% on the Dostour dataset. In general, classification of positive and negative texts has a much better F-score than that of neutral texts. The fact that these constituted a small percentage of the used datasets meant that accuracy was not degraded too much. This however might not always be the case which requires further enhancements.

The results also show that the proposed algorithms perform significantly better on the twitter dataset. Further analysis has revealed that this is largely due to the nature of the Dostour dataset itself in which many comments include suggested modifications to articles of the constitution which themselves contain sentiment terms but that in this context do not actually reflect a sentiment made by the author. Questions on articles caused confusion as well. An example in the data set is "ما الفرق بين التكافل والتضامن؟" the translation of which is "What is the difference between interdependence and solidarity?" This resulted in a positive classification when in fact this comment should have been classified as neutral. Another reason for the degraded performance was the fact that many sentiment terms used in the Dostour dataset were missing from the used lexicon suggesting that it needs to be further expanded.

VI. CONCLUSION

This paper highlighted problems and challenges that face researchers aiming to carry out sentiment analysis of Arabic social media. The work presented a case study that investigates the extent to which these problems can be overcome when determining the semantic orientation of microblogs. Preliminarily experiments show that the adopted approach is effective especially within the context of twitter despite the very limited resources used. Further modifications and experimentation on other datasets will be carried out to further improve and validate the proposed approach.

REFERENCES

[1] Semiocast, Geolocation analysis of Twitter accounts and tweets by Semiocast, 2012. [Online]. Available: http://semiocast.com/publications/2012_07_30_Twitter_reaches_half_a_billion_accounts_140m_in_the_US.

[2] "Facebook Statistics by Country," 2012. [Online]. Available: http://www.socialbakers.com/facebook-statistics/.

[3] S. Choney, "Middle East has the most politically vocal social networkers, says Pew," NBCNews, 2012. [Online]. Available: http://www.nbcnews.com/technology/technolog/middle-east-has-most-politically-vocal-social-networkers-says-pew-1C7559554.

[4] M. Taboada, M. Tofiloski, M. Stede, J. Brooke, and K. Vollz, "Lexicon-Based Methods for Sentiment Analysis," Computational Linguistics, vol. 1, no. 1, pp. 1–42, 2011.

[5] M. Elhawary and M. Elfeky, "Mining Arabic Business Reviews," 2010 IEEE International Conference on Data Mining Workshops, pp. 1108–1113, Dec. 2010.

[6] M. Abdul-Mageed, "Subjectivity and sentiment analysis of Modern Standard Arabic," in Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, 2011, no. 3, pp. 587–591.

[7] M. Abdul-Mageed and M. Diab, "Toward Building a Large-Scale Arabic Sentiment Lexicon," in Proceedings of the 6th International Global WordNet Conference, 2012, pp. 18–22.

[8] M. Abdul-Mageed, S. Kubler, and M. Diab, "SAMAR: A System for Subjectivity and Sentiment Analysis of Arabic Social Media," in Proceedings of the 3rd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis, 2012, pp. 19–28.

[9] A. Shoukry and A. Rafea, "Preprocessing Egyptian Dialect Tweets for Sentiment Mining," in Proceedings of the fourth workshop on Computational Approaches to Arabic Script-Based Languages, 2012, pp. 47–56.

[10] M. Sawalha and E. Atwell, "Fine-Grain Morphological Analyzer and Part-of-Speech Tagger for Arabic Text," in Proceedings of LREC'10, 2010, pp. 1258–1265.

[11] N. Habash, R. Eskander, and A. Hawwari, "A Morphological Analyzer for Egyptian Arabic," in Proceedings of the Twelfth Meeting of the Special Interest Group on Computational Morphology and Phonology (SIGMORPHON2012), 2012, pp. 1–9.

[12] P. Turney, "Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews," in Proceedings of the 40th Annual Meeting on Computational Linguistics (ACL), 2002, no. July, pp. 417–424.

[13] A. Esuli and F. Sebastiani, "Determining the semantic orientation of terms through gloss classification," in Proceedings of the 14th ACM international conference on Information and knowledge management, 2005, pp. 617–624.

[14] O. Zayed, S. R. El-Beltagy, and O. Haggag, "A Novel Approach for Detecting Arabic Persons' Names using Limited Resources," accepted for presentation at CICLing 2013.