COMPUSOFT, An international journal of advanced computer technology, 7(11), November-2018 (Volume-VII, Issue-XI)

Available online at: https://ijact.in

Date of Submission	18/10/2018
Date of Acceptance	16/11/2018
Date of Publication	30/11/2018
Page numbers	2884-2891 (8 Pages)

ISSN:2320-0790

An International Journal of Advanced Computer Technology

AN EXPANDABLE AND UP-TO-DATE LEXICON FOR SENTIMENT ANALYSIS OF ARABIC TWEETS

B.Ihnaini[1], M. Mahmuddin[2]

[1, 2]School of Computing, College of Art & Sciences,
Universiti Utara Malaysia (UUM), Sintok, 06010, Kedah, Malaysia
[1]baha_najim@ahsgs.uum.edu.my, [2]ady@ahsgs.uum.edu.my

Abstract: Sentiment analysis is the process of identifying the subjective opinion within a text. And it gains a huge interest due to its several benefits in developing economy, politic, and sociology. And since twitter is considered a rich source of people's thoughts and opinions, it is urged to benefit from it to explore public opinions. Many researches have been conducted for English language, while Arabic language still got limited number of sentiment analysis studies, especially in the context of Arab dialects in social media. A lexicon-based approach is adopted to perform sentiment analysis on Arabic tweets, which rely on detecting sentiment words. These sentiment words are loaded in a sentiment lexicon where words are annotated by its sentiment polarity. One of the main issues of handling Arabic tweets is the changing nature of twitter, where new words that imply sentiment values emerged, and many slang words are evolved. In this paper, an expandable and up-to-date lexicon for Arabic (EULA) is developed to overcome the issue of inventing new words and phrases in social media. EULA rely on a pre-built lexicon of MSA sentiment words, and a set of rules to expand and enrich it with dialectical polarity words from a small amount of labelled tweets, and a large amount of unlabeled tweets. For evaluation, eight different corpuses of Arabic tweets were selected. And a pre-processing phase that includes normalization and stemming is implemented to reduce the number of unique words to be analyzed for sentiment analysis. Experiments show that EULA improved the lexicon-based approach`s accuracy and F-1 score by more than 20% on average.

Keywords: Sentiment analysis; Lexicon-based approach; Social media; Modern Standard Arabic (MSA)

I. INTRODUCTION

Opinions and sentiment of people are expressed using a diverse language. This is seen in the well-formed data, like news and reviews, and data from social media. However, social media streams use informal communication, where misspellings and abbreviations exist. And people on these media streams are willing to be entertaining and funny. Moreover, these platforms lead to the emerging of new words and phrases due to the continuous change of subjects. Analyzing sentiment is affected by this changing nature of language along with the dynamic nature of social media. And when dealing with data from Twitter, one of the main issues that arise when dealing with tweets is also the emergence of new words that imply sentiment values, and the time-changing nature of Twitter [1]. No need to wait months to see language change on Twitter, changes can be detected within a single day [2]. And in another research, authors stated that topics change rapidly on social media, and people are inventing new words and phrases [3]. According to [4] topics and vocabulary change over time on Twitter. And many slang words on social media evolves over time [5]. Therefore, performance of SA models drops dramatically when applied on tweets collected at a later stage. Therefore, building an expandable and up-to-date lexicon is the second objective in this research.

Due to this change and flexibility of language on social media and the new terms that keep coming up, models trained on social media data may degrade over time. Many studies have used the existing natural language processing tool like information extraction (IE) tools [6]and syntactic parsers[7], yet, these tools are not available for many languages like Arabic. Any analyzing method in social media platforms should adapt to (a) different low-resource languages, (b) the vibrant nature of social media, and (c) working with no supervision. In this paper, an approach for bootstrapping EULA through sentimental clues from Arabic tweets is presented. The proposed approach is:

- Deals with the informality, creativity and the dynamic nature of social platforms.
- Is not dependent on language-dependent tools.
- Scales and fits many Arabic dialects in social media.

Expanding EULA relies on two main assumptions. First, terms with similar orientation sentiment tend to co-occur at the tweet level[8]. Second, terms of opposite orientation sentiment do not co-occur at the tweet level [9].

II. RELATED WORK

The efficiency of lexicon-based approach is determined mainly by the sentiment lexicons. Many methods implemented to generate such lexicons. Some researchers built lexicon manually [10], some used synonyms and antonyms to expand a small set[11], [12]. While others used methods based on statistics and semantics[13], such as PMI [14], and the links between the terms[15]. Authors of [10]have constructed manually a 200 words lexicon. The results have been affected by the size. They found that the stemmer is useful to reduce the size of the lexicon since multiple words have the same root at the lexicon. Also, authors of [16] created their 3,982 adjectives polarity lexicon manually. And by translating the seed list of [17]into Arabic, and using it along with some random words including objective words also, [18]expanded the list to obtain 150,000 words in their lexicon.

In another work, researchers started with 300 words as a seed from SentiStrength. Then these words translated to Arabic. Two individuals annotate every word by 1 for positive word, and -1 for negative words. Then using Sakhr dictionary, a synonyms list extended the lexicon. After adding emoticons to the lexicon, it ended with 2,376 entries [11]. While ArSeLex created manually by collecting 400 seed words then expanded manually by adding synonyms and antonyms, then automatically expanded from several online sources to form a 5,244 adjectives lexicon[19].

Building a sentiment lexicon was proposed by[20]in three steps: a) To collect Arabic stems. b) Translating these stems into English. c) Finding the sentiment score of these stems from English sentiment lexicons located online. Around 120,000 stems have been gathered into the lexicon.

A lexicon-based method was adopted in [21], they built a lexicon form extending a seed of 380 words. Then they annotate the lexicon`s words by using two algorithms, to present the earliest lexicon with scores to its entries. Finally, they implement two approaches to find the overall sentiment of the collected tweets: a) is called the sum, which sums the words` polarities in a tweet, and b) called the double polarity to find for each word in a tweet its positive and negative score. These two scores are given to every polarity word per the word’s frequency in the corpus.

The lexicon-based approach has been under trial for improving it in[22]. The construction of their lexicon consisted of four stages. a) The SentiStrength website was used for selecting 300 seed words. b) They added the synonyms of these words to the lexicon. c) A scheme called term frequency weighting was used for identifying the missed words in the preceding steps, and d) words from different Arabic dialects were added to the lexicon for enlargement. Then, a SA tool was created for calculating the text’s polarity without neglecting negation and intensification. Simple lexicon-based method was used, which depends on classifying the sentence per the higher number of negative or positive words counted in it.

Three SA methods for Arabic were presented in[23], a simple lexicon-based method was one of them, but its performance is enhanced by adding a set of features to handle valence shifters such as negation and intensification.

A comparison in performance of supervised and lexicon-based approaches have been held in [24]. The authors studied corpus-based and lexicon-based approaches for Arabic SA. They developed an Arabic lexicon using a seed of 300 words and then synonyms were added to the lexicon. And aggregated all polarity weights of these words after applying negation and intensification to these weights. They indicated that the performance is poor when the lexicon is not sufficiently large in the lexicon-based approach.

SANA was presented in [25], an Arabic sentiment lexicon. It combines manually created lexicons such as HUDA and SIFAAT, and involves manual annotations, automatic machine translation and gloss matching by using several resources, such as SAMA and THARWA. SANA includes around 225,000 entries, but many of them are inflected, duplicates or not diacritized, which makes it noisy and hard to be useable. It covers two dialects, Egyptian and Levantine and is not applied to SA tasks yet.

Authors of [26] built a large-scale sentiment lexicon for MSA, and call it ArSenL. ArSenL consist of 28,780 lemmas and 157,969 related synsets. It is a combination of four existing resources: English WordNet (EWN), ArabicWordNet[27], SAMA[28]and SentiWordNet [29]. But it does not include dialect words, only MSA. Therefore, the accuracy is affected when it is applied on social media.

Following the example of ArSenL, [15] constructed Standard Arabic Sentiment Lexicon (SLSA) by developing a matching algorithm between entries in SentiWordNet and entries of an Arabic morphological analyzer. A link is created then if there is a match between the entries from both. And the score is assigned to the entries from SentiWordNet. Nevertheless, SLSA just like ArSenL has no dialect words. Hence, social media text cannot be accurately analyzed.

Three lexicons for Arabic were generated from Twitter in [30]. Each one from a Twitter dataset collected as following: the first dataset contained emoticons, the second contained a seed list of hashtags for positive and negative Arabic words, and the third also contained hashtags of dialectical Arabic positive and negative words. Then three lexicons were generated using PMI from these datasets: 21,964 for Arabic Hashtag Lexicon (AHL), 20,128 for Dialectal Arabic Hashtag Lexicon (DAHL) and 43,304 for Arabic Emoticon Lexicon (AEL).

While authors of [31] built an Arabic lexicon in three steps: first, from [21] they used a learning algorithm that employs seed words to expand the lexicon. In the second phase, they used a lexicon created by [26]. This lexicon contains 154k words along with their punctuation. Since every punctuation make different meaning for the words, there were different punctuations for same word in the lexicon. But the usage of punctuations by tweeters is seldom; therefore, they removed them from the words. Then they removed Arabic diacritics from words. Because the same word could appear in the lexicon with different diacritics, therefore, the lexicon might have duplicate text. So, they removed duplicate entries from the lexicon. In the third phase, they added manually new words. And they finally get a large-scale lexicon, and it contains over 14k sentiment words.

Authors of [13]started from the earlier work of [21], and then assign scores to the Arabic words of the lexicon in three steps. First, by collecting 100 tweets for each word using Twitter's search API, ending with almost 500k tweets. Second, finding co-occurrence statistics from the dataset. Finally, find the score for each word using these statistics based on the hypothesis that the stronger a polar term is, the less likely it is to co-occur with terms of an opposite polarity.

Finally, [32] created a lexicon for the Algerian dialect only. they relied on the lexicon of [21], which its words from Egyptian dialect and MSA. First, they remove all words not used in the Algerian dialect. After that, they added all the words of the Algerian dialect which are equivalent to Egyptian and Arabic words in the lexicon. Finally, they added the commonly used words of the Algerian dialect that carry positive or negative opinion. At the end of these steps, the constructed lexicon ended with 3,093 polarity words.

III. DATASETS USED

In this paper, several datasets have been used for expansion EULA and evaluating the proposed system, all of them consist of Arabic tweets. Most of these datasets are imbalanced in favor of negative tweets. From the work of [33], ArSenTD-LEV was built, this Arabic Sentiment Twitter Dataset for the Levantine dialect consists of 4,000 Tweets divided equally among the four Levantine countries (Jordan, Lebanon, Palestine and Syria). The distributions of the tweets are 1,232 positive tweets, 1,883 negative tweets, and 885 neutral tweets. And it covers the common topics (politics, religion, sports, personal and entertainment), in this research only the 3,115 positive and negative tweets are considered since the desired classification is binary.

Another dataset called ASTD, that was built by [34], which is the Arabic Sentiment Tweets Dataset with approximately 10,000 tweets c1assified as 1,684 negative tweets, 799 as positive tweets, 832 as neutral and 6,691 as objective tweets. Only the 2,483 positive and negative tweets is considered in this research. In addition, the dataset RR was created by [35], consists of manually annotated 6,514 Arabic Tweets for sentiment. However, most of these are objective tweets, only 878 positive tweets and 1,943 negative tweets are considered in this research.

Moreover, TDS, a Twitter dataset was built by [21]through collecting 500 random tweets and manually annotating them with their semantic orientation. The 310 tweets classified as negative and the 155 tweets classified as positive were considered for this research, while the 35 neutral tweets were neglected. The JOR-Tweets is a balanced dataset that consists of 2,000 tweets were created by [24]. It covers various topics, such as politics and arts. And it is written in both MSA and Jordanian dialect. And another dataset which called Syrian-Tweets is consist of 2,000 tweets, among these tweets, 448 positive tweets and 1,350 negative tweets are used in this research. [14] Collected these tweets originating from Syria, which is another country from the Levantine countries.

And finally, authors of [33] created AraSenti-Tweet, a dataset that consists of 10,133 tweets. 4,329 tweets are positive, and the rest 5,804 tweets are negative. Table 1 shows the basic statistics of all of these datasets.

TABLE 1: DATASETS

Work	*Dataset*	*Positive*	*Negative*	*Total*	*Dialect*
[36]	ArSenTD-LEV	1,232	1,883	3,115	Levantine
[34]	ASTD	799	1,684	2,483	Egyptian
[35]	RR	878	1,943	2,821	MSA/ Dialectical
[21]	TDS	155	310	465	Egyptian
[24]	JOR-Tweets	1,000	1,000	2,000	MSA/ Jordanian
[14]	Syrian-Tweets	448	1,350	1,798	Syrian
[33]	AraSenti-Tweet	4,329	5,804	10,133	Saudi

IV. SYSTEM OVERVIEW

The proposed system requires several components. First, pre-processing the tweets. Then, since EULA requires a prebuilt lexicon, several lexicons have been collected, and another lexicon were constructed for this purpose, and it called the Basic lexicon, it consists of MSA sentimental words, with polarities in the range [-5, +5]. Third, create a set of rules to construct EULA.

A. Pre-processing Tweets

In this phase, five main tools are used to pre-process Arabic tweets[37]. First, the cleaning phase, such as removing re-tweets, removing URLs, removing user mentions in tweets, removing non Arabic letters, removing Twitter special characters, removing diacritics, removing punctuations and removing numbers.

Second, tokenization phase, by dividing the sentence to its terms. Then, stop words removal, by removing words that don`t add a value to the tweets, such as "من /of","على /on", "في /in".

Fourth, normalization phase, which is replacing many forms of a letter to one form. For example, (آ،أ،إ،ا) will be replaced with (ا). Finally, stemming phase, which is about removing the added suffixes that are used for different objectives, whether before word, or at its end. For example; in the word"المبدعون/ Creative" the suffix "ال" in the beginning, and "ون" at the end referring to plural, are both not necessary for the word's meaning and does not change its sentiment polarity. So after the stemming is applied on this word it will become "مبدع".

B. Sentiment Lexicons and Basic Lexicon

First of all, EULA rely on a prebuilt lexicon in order to find polarity words in a tweet, and then using the presence of the polarity words, negation words, and contrast words, it will be expanded and updated. From the literature, severeal lexicons have been built for Arabic. In this paper, ten public lexicons from Section 2 have been collected and used in the evaluation process. Besides, a Basic lexicon is built in this paper.

First, the UnWeighted Opinion Mining (UWOM) Lexicon, which is the work of [21], where they had built their first lexicon that consists of 4,392 terms. It includes MSA terms and Egyptian terms with binary polarity +1 or -1. Second, ArSenl, A large-scale MSA sentiment lexicon, which was built by [26]. It consists of 28,780 lemmas and 157,969 related synsets, but after removing the diacritics, since Arab writers rarely use them on social media, it ends up with 34,821 words, with weights ranging from -1 to +1.

Third, authors of [15]constructed SLSA. It is composed of around 35,000 entries annotated with scores ranging from -1 to +1. In addition, NileULex lexicon, which have been built in [38]. It contains 5,953 unique terms. Almost half of these terms are MSA, and the others are Egyptian. All terms are assigned by binary polarity.

Authors of [14] created three different lexicons. First, DAHL, that have been generated from a dataset contained seed list of hashtags for positive and negative dialectical Arabic words, with weights that varies from -5.8 to +5.3. And it ends with 20,128 sentiment words. Second, AHL, using dataset formed by a seed list of positive and negative Arabic words, and it consists of 21,964 sentiment words. And the weights varies from -8.4 to +4.7. And finally AEL, where a dataset contained emoticons is used to generate this lexicon, and it ends up with 43,304 sentiment words. And the scores of these words are in the range -4.5 to +4.0.

While authors of [23] created two lexicons. First, AraSenti-PMI lexicon which have been constructed through finding the PMI measure of all the words in a positive and negative tweets, the lexicon consists of 56,938 positive terms, and 37,023 negative term, with weights ranging from -7 to +5.5. Second,AraSenti-Trans lexicon, it is a binary polarity lexicon that consists of 59,525 positive words, and 71,817 negative words, by utilizing two popular sentiment lexicons in English, the MPQA lexicon [39], and the Liu lexicon[40].

Moreover, in the work of [41], Arabic Senti-Lexicon were built by two methods, manually by three annotators, and automatically based on the appearance of the sentiment terms in positive and negative reviews. It ends with 3,948 multi dialects polarity words, with scores ranging from -5 to +5.Table 2 shows a summary of these lexicons.

TABLE 2: LEXICONS

Work	*Lexicon*	*Size*	*Polarity*	*Dialect*
[21]	UWOM Lexicon	4,392	Binary	Egyptian
[26]	ArSenL	34,821	Scores	MSA
[15]	SLSA	35,000	Scores	MSA
[38]	NileULex	5,953	Binary	Egyptian
[14]	DAHL	20,128	Scores	Multi-dialect
[14]	AHL	21,964	Scores	MSA
[14]	AEL	43,304	Scores	Multi-dialect
[23]	AraSenti-PMI	93,961	Scores	Saudi
[23]	AraSenti-Trans	131,342	Binary	Saudi
[41]	Arabic Senti-Lexicon	3,948	Scores	Multi-dialect
Author`s work	Basic Lexicon	4,266	Scores	MSA

Finally, the Basic lexicon consist of MSA sentiment words that have been translated from well-known pre-built English lexicons, namely SentiStrength consists of around 2,550 polarity words, AFINN consists of 2,460 polarity words, VADER consists of more than 7,500 polarity words, and finally MPQA which consists of around 8,200 polarity words. The weight score of these words are adjusted to range score of [-5, +5]. Then 15 simple lexicon-based experiments were conducted to find the best combination between these 4 English lexicons. And results show that the combination of SentiStrength, VADER, and MPQA produced the highest accuracy as presented in Table 3. Finally, the obtained Basic lexicon consists of 4,266sentiment words.

TABLE 3: FINDING BEST COMBINATION FOR THE BASIC LEXICON

Lexicons combination	*Size*	*Accuracy*
AFINN	1,955	40.0%
SentiStrength	1,617	36.0%
VADER	2,439	42.0%
MPQA	3,465	45.8%
AFINN+ SentiStrength	2,605	46.0%
AFINN+ VADER	2,454	45.9%
AFINN+ MPQA	3,910	50.2%
SentiStrength+ VADER	2,859	47.9%
SentiStrength+ MPQA	3,782	47.3%

VADER+ MPQA	4,090	52.3%
AFINN+SentiStrength+ VADER	2,869	47.4%
AFINN+ SentiStrength+ MPQA	4,147	50.9%
AFINN+ VADER+ MPQA	4,091	51.1%
SentiStrength+ VADER+ MPQA	4,266	**52.5%**
AFINN+ SentiStrength+ VADER+ MPQA	4,276	51.5%

C. EULA Expansion

Two types of resources to expand EULA and make it up-to-date, through unlabeled and labeled tweets. The labeled tweets are used to predict the polarity of the unknown words, taking into consideration the presence of valence shifter words, next section explains in detail how to perform it. And the unlabeled tweets are used in several ways to predict the polarity of unknown words, such as using contrasts word and some predictor words to predict the polarity of unknown words, furthermore, using the assumption of that positive words tend to appear with other positive words, and negative words tend to appear with other negative words. These predicted word`s polarity will be sent to EULA, where there is a counter for positive occurrences, a counter for negative occurrences, a sum of the previous counters, and the polarity for each word which is represented in the following equation:

Polarity(X) = 5*(Cp- Cn)/Cs (1)

Where: *Cp* represents the number of times *X* is added to EULA as positive, *Cn* represent the number of times *X* is added to EULA as negative, and *Cs* represent the summation of *Cp* and *Cn*. For example, if *X* have been added as positive 9 times, and 1 time as negative, the polarity of *X* will be assumed as: 5*(9-1)/(10), which equal +4, a high positive weight that fit in the proposed system that consider weights from -5 to +5. There are two ways to expand EULA, the first is through labelled tweets, and the second through unlabeled tweets. Next sub-sections explain in detail each one.

1) Expanding EULA through labeled tweets

First, labeled tweets are used to add words with its weights to EULA as following, if word *w* is found in a tweet, but word *w* does not exist in the prebuilt lexicon, the unknown word*w* will be assigned positive if the tweet is annotated positive, and will be assigned negative if the tweet is negative. Unless a negation or contrast word appears in the tweet, then, the word in the scope of negation will be assigned opposite of the tweet's polarity. Moreover, if contrast appears in the tweet, two different actions will be applied depending on the type of the contrast word; if it was *C1*, where the main sentence is before it, the unknown words before the contrast word will have the same polarity of the tweet`s annotation, and the unknown words after the contrast word will have the opposite polarity of the tweet`s annotation. And if it was *C2*, where the main sentence is after it, then the polarity of words before contrast word are the opposite to the polarity of the tweet, and after it is same as the tweet`s polarity. In the end, the unknown word along with its annotation will be sent to EULA as shown in Algorithm 1.

Algorithm 1: Expanding EULA through labeled tweets

1 Input: Tweet as T, and labeled as L
2 Output: predicted strength of new terms sent to EULA
3 For each unknown word U in T
4 If Contrast of type C1 ϵ T
If U is before C1:
then U is labeled as L
If U is after C1:
then U isopposite of L
5 If Contrast of type C2 ϵ T:
If U is before C2:
then U is opposite of L
If U is after C2:
then U is labeled as L
6 If Negation word ϵ T:
If U in scope of Negation:
then U is opposite of L
7 Else
U is labeled as L
8 Send U with its annotation to EULA
9 End For

2) Expanding EULA through unlabeled tweets

In the second way in expanding EULA, unlabeled tweets are used, as presented in Algorithm 2, which will function as follows; the first source is that if the unknown word appears in a tweet that has words with only one type of polarities, either positive or negative beside non polarity words, then the unknown word will be assigned to positive or negative based on this polarity and sent to EULA, unless a negation word exist in the tweet, where the polarity is flipped if the polarity word or the unknown word is in the scopeof the negation word. And if the words in the tweet have a mix of both polarities of positive and negative; then the unknown word will not be sent to EULA.

Algorithm 2: Expanding EULA through unlabeled tweets

1 Input: unlabeled tweet as T
2 Output: predicted strength of new terms sent to EULA
3 If Contrast ϵ T
Let text before Contrast is T1
Let text after Contrast is T2
If positive word ϵ T1 AND
negative words ∉ T, then:
Use T2 as labeled negative tweet
If negative word ϵ T1, AND
positive words ∉ T, then:
Use T2 as labeled positive tweet
If positive word ϵ T2 AND
negative words ∉ T, then:
Use T1 as labeled negative tweet
If negative word ϵ T2, AND
positive words ∉ T, then:
Use T1 as labeled positive tweet
4 If predictor word type 1 ϵ T, then:
Send next unknown word as positive to EULA
5 If predictor word type 2 ϵ T
Send next unknown word as negative to EULA
6 If positive words ϵ T AND
negative words ∉ T, then:
Unknown words is sent to EULA

```
    as positive
If negative words ϵ T AND
  positive words ∉ T, then:
    Unknown words is sent to EULA
    as negative
End
```

V. Results and Discussion

In order to evaluate EULA, prebuilt lexicons from Section 4.2, datasets from Section 3 are used for evaluation purposes, and by impleminting simple lexicon-based several experiments is conducted. First, the performance of the prebuilt lexicons is found individually by implementing a simple lexicon-based approach test without using EULA, as in Figure 1. Then, EULA is expanded by labeled tweets, along with each lexicon on all of the datasets, as in Figure 2. And finally, as in Figure 3, EULA is expanded by unlabeled tweets, along with each lexicon on all datasets. Then, all the performance metrics are found for each experiment.

First, each datasetis splitted into 2 datasets, one for expanding EULA which is called training dataset, and another one for evaluation purposes called testing dataset. Second, for each experiment one of the eleven prebuilt lexicons is used.

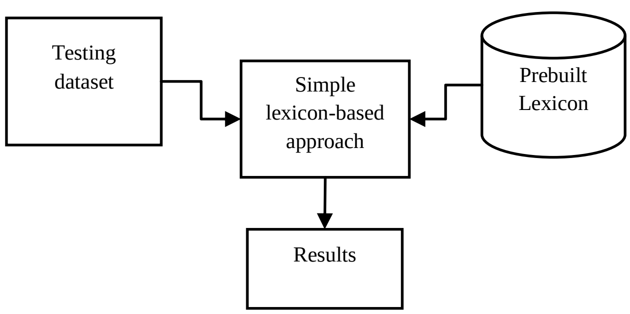

Figure 1: Experiment layout without EULA

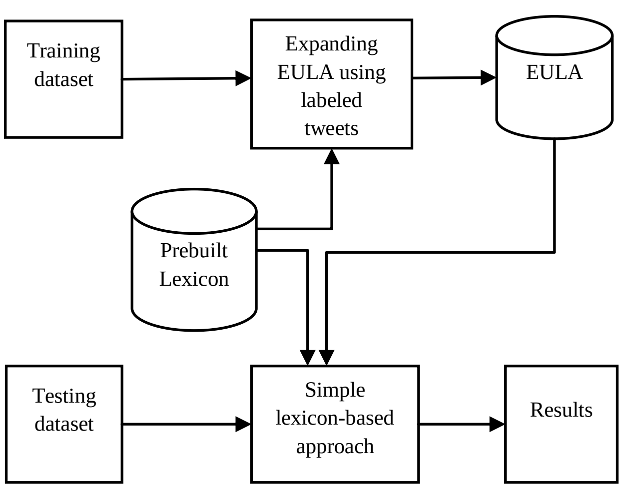

Figure 2: Experiment layout with EULA expanded by unlabeled tweets

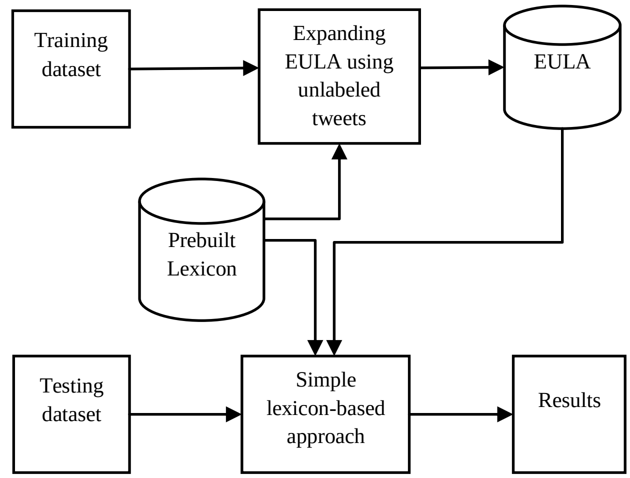

Figure 3: Experiment layout with EULA expanded by labeled tweets

For the evaluation process in sentiment analysis the following metrics are the most famous ones, which are accuracy, precision, recall and F-measure score [42]. And to explain them, first the TP, FP, TN, and FN shall be explained.

- True positives (TP) shows the division of correct classification for instances of positive class by total number of instances. These positive instances are those found positive by the system and also positive in real.
- False positives (FP) shows the division of the incorrect classification of positive instances by total number of instances. These instances are found positive, but in real it is negative.
- True negatives (TN) shows the division of correct classification for instances of negative class by total number of instances. These instances that are found negative by the system and also negative in the real.
- False negatives (FN) shows the division of the incorrect classification of negative instances by total number of instances. These instances are found negative, but in real it is positive.

The accuracy metric represents the number of instances that have a correct classification among all classes:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FP} \quad (2)$$

Precision shows the portion of the recovered relevant instances; the number of positive-classified instances that are positive in reality:

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

Recall represents the portion of the relevant instances which are recovered; the number of instances that are in reality positive to have a positive classification:

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

The F-measure score is a metric of accuracy that combines precision and recall as:

$$F1 = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

Accuracy, precision, recall and F-measure scores are found as an average values for each experiment, then it will be compared to each other.Table 4, presents the precision and recall for each lexicon without using EULA, with EULA expanded through unlabeled tweets, and with EULA expanded through labeled tweets. And results show that using EULA through unlabeled tweets improved the precision and recall by 11% and 8% respectively. While using EULA through labeled tweets improved the precision and recall by 25% and 22% respectively. Therefore, it is clear that using EULA improves the lexicon-based approach in term of precision and recall by a good margin, especially when using labeled tweets to expand bootstrap EULA. And it can be noticed that EULA improve the performance when tested along with MSA sentiment lexicons more than when combined with a dialectical lexicon.

TABLE 4: PRECISION AND RECALL, WITH AND WITHOUT USING EULA

Lexicon	*Average without EULA*		*Average with EULA using unlabeled tweets*		*Average with EULA using labeled tweets*	
	P	**R**	**P**	**R**	**P**	**R**
AEL	42.5	42.8	45.8	46.0	68.4	67.1
AHL	55.2	54.9	64.5	55.8	72.1	70.0
Arabic Senti-Lexicon	31.9	32.0	56.1	52.8	72.1	69.0
AraSenti-PMI	68.6	68.2	66.2	64.1	72.5	70.2
AraSenti-Trans	59.1	58.8	58.2	54.3	69.6	66.1
ArSenL	27.6	27.8	46.4	47.2	69.9	67.9
Basic-Lexicon	52.0	51.9	64.6	64.1	71.5	69.8
DAHL	55.1	55.0	59.2	57.5	71.7	70.1
NileULex	38.6	38.6	57.8	51.9	70.0	66.5
SLSA	40.0	39.9	54.5	51.8	69.9	67.2
UWOM Lexicon	35.3	35.2	57.3	51.9	70.1	66.8
Average	46.0	45.9	57.3	54.3	70.7	68.2

Note: P is Precision, R is Recall

Table 5 presents the accuracy and F-measurement score that have been found in the previous experiments. And it shows that expanding EULA using unlabeled tweets improved the accuracy and F-measurement score by 7% and 10% on average respectively. While expanding EULA through labeled tweets increased average of the accuracy and F-measurement score by 21% and 22% respectively. And from these results, it is concluded that EULA contributed in a big improvement to the lexicon-based approach. Especially, when expanded by a sufficient amount of labeled tweets.

Besides, the improvement is better in case of MSA lexicons, such as in ArSenl, and SLSA, where the improvement were 40% and 27% respectively.

TABLE V: ACCURACY AND F-MEASUREMENT SCORE WITH AND WITHOUT USING EULA

Lexicon	*Average without EULA*		*Average with EULA using unlabeled tweets*		*Average with EULA using labeled tweets*	
	A	**F1**	**A**	**F1**	**A**	**F1**
AEL	42.8	42.2	46	45.90	67.1	66.6
AHL	54.9	53.9	55.8	59.84	70.0	69.3
Arabic Senti-Lexicon	32.0	31.8	52.8	54.40	69.0	67.7
AraSenti-PMI	68.2	68	64.1	65.13	70.2	69.3
AraSenti-Trans	58.8	58.4	54.3	56.18	66.1	64.4
ArSenL	27.8	27.6	47.2	46.80	67.9	67.1
Basic-Lexicon	52.5	51.9	64.1	64.35	69.8	69.2
DAHL	55.0	54.7	57.5	58.34	70.1	69.5
NileULex	38.6	38.4	51.9	54.69	66.5	64.9
SLSA	39.9	39.8	51.8	53.12	67.2	66
UWOM Lexicon	35.2	34.9	51.9	54.47	66.8	65.2
Average	46.0	45.6	54.3	55.78	68.2	67.2

Note: A is Accuracy; F1 is the F-measurement score

VI. CONCLUSION

In this paper, EULA have been proposed to handle the emergence of new dialectical Arabic words in social media, by presenting a scalable and dialectical independent bootstrapping approach to expand and learn from Arabic tweets. Results show the effectiveness of this approach and the improvement it give to lexicon-based approach sentiment analysis.

The results confirm that the approach can be effectively exploited and further improved for subjectivity classification for many Arabic dialects in social media.

VII. REFERENCES

[1] J. Eisenstein, "What to do about bad language on the internet," Naacl-Hlt, pp. 359–369, 2013.

[2] S. Golder and M. Macy, "Diurnal and Seasonal Mood Vary with Work, Sleep, and Daylength Across Diverse Cultures," Science, vol. 339, no. February, pp. 819–822, 2013.

[3] S. Volkova, T. Wilson, and D. Yarowsky, "Exploring Sentiment in Social Media: Bootstrapping Subjectivity Clues from Multilingual Twitter Streams," Proc. 51st Annu. Meet. Assoc. Comput. Linguist. (Volume 2 Short Pap., pp. 505–510, 2013.

[4] E. Refaee and V. Rieser, "Benchmarking Machine Translated Sentiment Analysis for Arabic Tweets," Proc. NAACL-HLT 2015 Student Res. Work., pp. 71–78, 2015.

[5] H. Elsahar and S. R. El-Beltagy, "A fully automated approach for Arabic slang lexicon extraction from microblogs," Lect. Notes Comput. Sci., vol. 8403 LNCS, no. PART 1, pp. 79–91, 2014.

[6] E. Riloff, and J. Wiebe, "Learning extraction patterns for subjective expressions," Proc. 2003 Conf. Empir. methods Nat. Lang. Process. - , pp. 105–112, 2003.

[7] J. Wiebe, T. Wilson, and C. Cardie, "Annotating expressions of opinions and emotions in language," Lang. Resour. Eval., vol. 39, no. 2–3, pp. 165–210, 2005.

[8] P. Turney, "Thumbs up or thumbs down? Semantic Orientation applied to Unsupervised Classification of Reviews," Proc. 40th Annu. Meet. Assoc. Comput. Linguist., no. July, pp. 417–424, 2002.

[9] A. Aue and M. Gamon, “Customizing Sentiment Classifiers to New Domains: A Case Study.,” Proc. Recent Adv. Nat. Lang. Process., vol. 3, no. 3, pp. 16–18, 2005.

[10] L. Albraheem and H. S. Al-Khalifa, “Exploring the problems of sentiment analysis in informal Arabic,” Proc. 14th Int. Conf. Inf. Integr. Web-based Appl. Serv. - IIWAS ’12, p. 415, 2012.

[11] R. M. Duwairi, N. A. Ahmed, and S. Y. Al-Rifai, “Detecting sentiment embedded in Arabic social media - A lexicon-based approach,” J. Intell. Fuzzy Syst., vol. 29, no. 1, pp. 107–117, 2015.

[12] H. S. Ibrahim, S. M. Abdou, and M. Gheith, “Idioms-proverbs lexicon for modern standard Arabic and colloquial sentiment analysis,” Int. J. Nat. Lang. Comput., vol. 4, no. 2, pp. 95–109, 2015.

[13] Samhaa R. El-Beltagy, “WeightedNileULex: A Scored Arabic Sentiment Lexicon for Improved Sentiment Analysis,” Lang. Process. Pattern Recognit. Intell. Syst. Spec. Issue Comput. Linguist. Speech& Image Process. Arab. Lang., no. February, 2017.

[14] S. M. Mohammad, M. Salameh, and S. Kiritchenko, “How translation alters sentiment,” J. Artif. Intell. Res., vol. 55, no. January, pp. 95–130, 2016.

[15] Shoeb, M., & Gupta, V. K. (2013). A crypt analysis of the tiny encryption algorithm in key generation. International Journal of Communication and Computer Technologies, 1(38).

[16] M. Abdul-mageed, M. Diab, and S. Kübler, “SAMAR : Subjectivity and sentiment analysis for Arabic social media,” Comput. Speech Lang., pp. 1–18, 2013.

[17] P. D. Turney and M. L. Littman, “Measuring Praise and Criticism: Inference of Semantic Orientation from Association,” ACM Trans. Inf. Syst., vol. 21, no. 4, pp. 315–346, 2003.

[18] H. Mobarz, M. Rashown, and I. Farag, “using automated lexical resources in arabic sentence subjectivity,” Int. J. Artif. Intell. Appl., vol. 5, no. 6, 2014.

[19] H. S. Ibrahim, S. M. Abdou, and M. Gheith, “sentiment analysis for modern standard arabic and colloquial,” 2015 IEEE 2nd Int. Conf. Recent Trends Inf. Syst. ReTIS 2015 - Proc., vol. 4, no. 2, pp. 353–358, 2015.

[20] M. Al-ayyoub, S. B. Essa, and I. Alsmadi, “Lexicon-based sentiment analysis of Arabic tweets,” Int. J. Soc. Netw. Min., vol. 2, no. July 2016, pp. 101–114, 2015.

[21] S. El-Beltagy and A. Ali, “Open issues in the sentiment analysis of Arabic social media: A case study,” … Inf. Technol. (IIT), 2013 9th …, pp. 1–6, 2013.

[22] N. A. Abdulla, N. A. Ahmed, M. A. Shehab, M. Al-Ayyoub, M. N. Al-Kabi, and S. Al-rifai, “Towards Improving the Lexicon-Based Approach for Arabic Sentiment Analysis,” International Journal of Information Technology and Web Engineering, vol. 9, no. 3. pp. 55–71, 2014.

[23] shah, a., sanghvi, k., sureja, d., & seth, a. (2018). insilico drug design and molecular docking studies of some natural products as tyrosine kinase inhibitors. international journal of pharmaceutical research, 10(2).

[24] N. a Abdulla, N. a Ahmed, M. a Shehab, and M. Al-ayyoub, “Arabic Sentiment Analysis: Corpus-based and Lexicon-based,” Jordan Conf. Appl. Electr. Eng. Comput. Technol., vol. 6, no. 12, pp. 1–6, 2013.

[25] M. Abdul-Mageed and M. Diab, “SANA: A Large Scale Multi-Genre, Multi-Dialect Lexicon for Arabic Subjectivity and Sentiment Analysis,” Proc. Lang. Resour. Eval. Conf., pp. 1162–1169, 2014.

[26] G. Badaro, R. Baly, and H. Hajj, “A Large Scale Arabic Sentiment Lexicon for Arabic Opinion Mining,” Arab. Nat. Lang. Process. Work. co-located with EMNLP 2014, Doha, Qatar, pp. 176–184, 2014.

[27] D. Vilares, C. Gómez-Rodríguez, and M. A. Alonso, “Universal, Unsupervised, Uncovered Sentiment Analysis,” no. July, 2016.

[28] K. Graff, D., Maamouri, M., Bouziri, B., Krouna, S. and T. S., and Buckwalter, “Standard Arabic Morphological Analyzer (SAMA) Version 3.1.,” Linguist. Data Consort. LDC2009E73., p. 2018, 2009.

[29] S. Baccianella, A. Esuli, and F. Sebastiani, “SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining,” Proc. Seventh Int. Conf. Lang. Resour. Eval., vol. 0, no. January, pp. 2200–2204, 2010.

[30] S. M. Mohammad, M. Salameh, and S. Kiritchenko, “Sentiment Lexicons for Arabic Social Media,” Tenth Int. Conf. Lang. Resour. Eval. Lr. 2016, no. September, pp. 33–37, 2016.

[31] A. Assiri, A. Emam, and H. Al-Dossari, “Towards enhancement of a lexicon-based approach for Saudi dialect sentiment analysis,” J. Inf. Sci., p. 016555151668814, 2017.

[32] M. Mataoui, “A Proposed Lexicon-Based Sentiment Analysis Approach for the Vernacular Algerian Arabic.,” Res. Comput. Sci., vol. 110, no. April 2016, pp. 55–70, 2016.

[33] N. Al-twairesh, “AraSenTi-Tweet : A Corpus for Arabic Sentiment Analysis of Saudi Tweets,” Procedia Comput. Sci., vol. 117, no. November, pp. 63–72, 2017.

[34] M. Nabil, “ASTD : Arabic Sentiment Tweets Dataset,” no. September, pp. 2515–2519, 2015.

[35] E. Refaee and V. Rieser, “An Arabic twitter corpus for subjectivity and sentiment analysis,” Proc. Lang. Resour. Eval. Conf., no. spring 2013, pp. 2268–2273, 2014.

[36] Munot, N. M., Lasure, P., & Girme, S. S. (n.d.). Design and Evaluation of Chronotropic Systems for Colon Targeted Drug Delivery. International Journal of Pharmacy Research and Technology (Vol. 2, pp. 13–17).

[37] B. Ihnaini and M. Mahmuddin, “An expandable and up-to-date lexicon for sentiment analysis of Arabic tweets,” J. Eng. Appl. Sci., vol. 13, no. 17, pp. 7313–7322, 2018.

[38] S. R. El-Beltagy, “NileULex: A Phrase and Word Level Sentiment Lexicon for Egyptian and Modern Standard Arabic,” to Appear Proc. Lr. 2016, no. April, pp. 2900–2905, 2016.

[39] T. Wilson, J. Wiebe, and P. Hoffman, “Recognizing contextual polarity in phrase level sentiment analysis,” Acl, vol. 7, no. 5, pp. 12–21, 2005.

[40] M. Hu and B. Liu, “Mining and summarizing customer reviews,” Proc. 2004 ACM SIGKDD Int. Conf. Knowl. Discov. data Min. KDD 04, vol. 04, p. 168, 2004.

[41] T. Al-Moslmi, M. Albared, A. Al-Shabi, N. Omar, and S. Abdullah, “Arabic senti-lexicon: Constructing publicly available language resources for Arabic sentiment analysis,” J. Inf. Sci., p. 0165551516683908, 2017.

[42] K. Schouten and F. Frasincar, “Survey on Aspect-Level Sentiment Analysis,” IEEE Trans. Knowl. Data Eng., vol. 28, no. 3, pp. 813–830, 2016.