Preprints.org

Article

Arabic X(formerly Twitter) Sentiment Analysis using Skip grams method

Rasha M. AlEidan [*]

Article

Arabic X (Formerly Twitter) Sentiment Analysis using Skip Grams Method

Rasha AlEidan

Collage of Computer Science, King saud University, Saudi Arabia, Riyadh; raleidan@ksu.edu.sa

Abstract: Nowadays and in the web era, there is a huge amount of information available due to automatically and easily generated websites. This information has many shared opinions from people toward different entities. Over the last decade there was a significant huge interest to determine the sentiment of any type of text especially on twitter. Twitter provides unstructured text that has many misspelling words, slang words and shortened forms. More complexity and challenges have come to the sentiment analysis researchers in the context of twitter. In this paper we conducted a literature review of the recent research of Twitter sentiment analysis on different languages. After that, we investigated challenges related to twitter sentiment analysis of Arabic language. Then we found a new solution for Sentiment analysis of Arabic twitter. So, we designed the main component of the proposed system and presented its algorithm. Finally, we performed many experiments under two kinds of datasets. The system achieves highest accuracy and shows many important results..

Keywords: Sentiment; Arabic; machine learning; Text analysis; Tweet; Social Network

1.Introduction

In our daily life, we receive and listen to information from different fields and sources. This information might be fact or opinions [1]. Opinions express the attitudes and sentiments of any entities in the real world: person, service or an event. The positive, negative or neutral are the main directions for people's opinions. Capturing the overall attitude is important in order to know the degree of quality for a product or movie as an example. This is not easy to acquire in real life because we have to hear and contact people in addition to reading their thoughts from news and media.

Nowadays and in the web era, there is a huge amount of information available and rapidly increasing due to automatically and easily generated websites. Consequently, any big or small organization, group of persons and individuals can share opinions in one place and online through the web. As a result of this, the sentiment analysis of web information is becoming an increasingly important area to research.

Sentiment analysis (or opinion mining) involves determining the evaluative nature of a piece of text [2]. Sentiment analysis aims to automatically extract and classify sentiments (the subjective part of an opinion) and/or emotions (the projections or display of a feeling) expressed in text [3]. Sentiment analysis has become a research field since the early 2000s [4]. Its impact can be seen in many practical applications, ranging from analyzing product reviews to predicting sales and stock markets using social media monitoring [4].

To the best of our knowledge, there are a few attempts for Arabic web content. Thus, there is an urgent need for automating the task of opinion mining of Arabic web content due to the expansion of the Arabic content in the last couple of years. Statistic website provides information on the languages in which the world leaders tweet of 2023 (see figure 1). English was found to be the most tweeted language among world leaders with 34% tweets. Japanese was ranked second, then, the third was Spanish with 12% and the Arabic language was fifth with 6% [5]. Therefore, our research will focus on sentiment analysis of Arabic text content published in a major micro-blogging service on the web, which is Twitter. Again, to the best of our knowledge, this work at this time is considered as the

first work aiming at exploring the sentiment of Twitter posts based on a set of features using skipgrams method and targeted for the Arabic language.
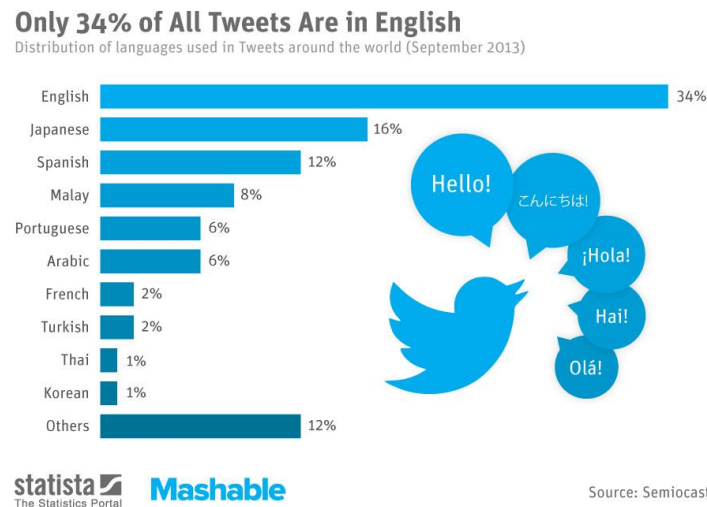

Figure 1. Percentage of tweet usage at 2023.

There are tools to evaluate the web information sentiments and to help users' to make decisions based on these viewpoints. Most of the researches are on the Spanish [5] [6], English [7] [2] and Chinese [8] websites. Unfortunately, there are little studies for Arabic web content. Therefore, it is crucial for the Arabic research community to build similar tools that assist in evaluating the sentiments of Arabic web content. In addition, the importance of the sentiment analysis research of social media in all languages is still in its initial stages and needs more efficient methods.

This research is important in the web era. People express their opinion in a short text through social networks. This is easy to use because of the existence of mobile phones that increases sharing thoughts through social networks. Such kind of data can be efficiently used for marketing, public opinion and user behavior analysis in addition to social study. Sentiment analysis of tweets text (as an example of social media) has several challenges. Tweets are limited in length, usually spanning one sentence or less. They tend to have many misspellings, slang terms, and shortened forms of words. They also have special markers such as hashtags [2]. In addition, this research is important to enhance Arabic foundation of online sentiment analysis especially for social media, as that field of research is very recent.

The key contributions of this paper are:

- Review of some previous works for automatic sentiment analysis on social media especially Twitter.
- Present the challenges and difficulties of twitter sentiment analysis in Arabic language.
- Propose an efficient Domain independent learning-based method to evaluate the opinions of Arabic text on Twitter based on skipgrams method.
- Perform many experiments to evaluate the performance of the proposed method on available Arabic tweet sentiment datasets.

2. Related Work

A significant huge interest in Twitter sentiment analysis research in different languages. The Twitter sentiment analysis studies can be divided into three categories based on their approach: Lexicon-based, learning-based and Machine translation–based (MT).

The Lexicon-based is depending on the existence of sentiment words with their polarities. This technique requires a huge effort from humans to build such lexicons that might be different from one domain to another. The big difficulty on twitter comes from online words and languages that are changing frequently depending on the slang of each environment or region.

For instance, [5] presented an unsupervised lexicon based method for sentiment analysis of Spanish twitter. They depended on three main lexicons for Opinion words, hashtags and emoticons in addition to a list of negation words. They introduced a simple heuristic method for building a hashtag lexicon by finding manually a list of positive and negative seed words. Then, any existing hashtags in the tweets that are retrieved by these seed words will be added to the lexicon with the seed word polarity. In addition to that, they provided a novel detection of negation words by using the dependency tree. The reported accuracy level of this approach without negation was 61% and 51% with negation. Furthermore, [7] proposed a hybrid system TOM framework that applies three different classifiers for English tweets in a pipeline way. First, the tweets are pre-processed via the suggested method. Then, the tweets are classified to positive and negative depending on the existence of the emotions. If the emotion is not found the tweets then will be considered as neutral. Secondly, the polarity classification, which is based on a set of positive and negative words, is applied.

This classifier classifies the neutral tweets by setting a score of each word and then sums the final polarity score of each tweet. Finally, the SentiWordNet classifier is applied on neutral tweets. If all the three techniques classify any tweet as neutral, they were considered as neutral. Their system achieves an average accuracy of 85.7% with 85.3% precision and 82.2% recall.

A domain dependent system was proposed by [8] for mining and sentiment analysis of Chinese twitter. They focused on the comments of hot tweets that are retweeted frequently by users. First and after preprocessing of tweets corpus, their method extracts the opinion targets from noun and noun phrases only. It uses an association-mining algorithm to find the hot opinion target words. The remaining terms of tweets, that are adjectives, adverbs and verbs and also contain the previous extracted opinion targets, are used to find a list of the opinion words. Then, the system generates the sentiment lexicon using SO-PMI, a measure of association, between a seed of manually labeled negative and positive words and new words from real tweet corpus. Finally the system applies a sentiment classification for each document on every target based on opinion words and sentiment lexicons.

As shown above, the advantage of previous lexicon-based methods is no need for labeled tweets as for supervised learning based methods because they depend on a series of score calculations.

For the second approach, the learning based systems have also been proposed to detect the sentiment of tweets. [2] proposed a supervised statistical sentiment analysis for short English informal texts such as tweets and SMS messages based on message and term levels. This method is important because it provides sentiment research with a lot of combinations of features that are tested with many experiments. Also by using many sentiment resources and data sets to compare between them. Their result obtains that linear SVM outperforms the maximum entropy classifier. In addition, the system is more efficient by using their new tweet specific lexicons at both types of classification (supervised or unsupervised).

For sentiment analysis of Spanish twitter [6] introduced a novel supervised leaning method. First the method does some preprocessing steps for the text. Then, it combines the tokenized terms to generate the skipgrams terms. This method is novel due to the use of skipgrams methods as a feature for SVM. It achieves a better performance than using single words and ngrams. Also no need for external knowledge and resources that are good to be used for Arabic language because of the lack of Arabic lexicons.

Additionally, [4] implemented a supervised sentiment analysis system for the Czech environment. They used a set of features that includes: N-grams, POS, Emotions and variant TFIDF features. They constructed a sentiment corpus of 10,000 posts from facebook and also more from product and movie reviews. Their system trains using two classifiers MXE and SVM. Finally, They obtained from their experiments on various set of features many important results.

In like manner, [9] provided a sentiment analysis system of Indian election. In detail, this work performs three analyses: Trend, Volume and Sentiment analysis in order to find the overall view of users to political parties. They used a naïve bays classifier for polarity classification of election tweets. In fact, this work only proposes a system without bringing the real experiments and results.

Regarding the third approach, the Machine Translation-based in the literature was used in two ways. The first way is partial Translation-based that used for the expansion of the non-English lexicons only. For example [10] presented a semi-supervised sentiment analysis for modern standard Arabic and Egyptian delicate including: tweets, product reviews, TV program comments and Hotel reservation data. Their work concerns with the problem of not standard Arabic -delicate language- that we can see its affect on most of their selected features. They used idioms/saying in their sentiment system including some delicate Egyptian. In addition to the upgrading of sentiment lexicons and some features resources to the delicate Egyptian such as the negation, intensifiers and wishful lists. This study based on sentiment lexicon from about (5244) annotated sentiment words only. Then they performed an expandable process by finding the synonym and antonym sets from free online Arabic dictionaries and lexicons to predict the sentiment of these extra words. They obtained high performance with accuracies of over 95%.

In the same line of thinking, [11] proposed unsupervised system analysis for Arabic twitter with the help of translation to build the lexicon. They translated the English lexicon of Sentistrengh system and also expanded the translated lexicon in Arabic using Arabic thesauri. Their system is simple and uses a little set of features (negation and Emoticons) and without considering the dialect language. At their experiment, they showed that stemming of Arabic tweet does not improve the overall accuracy of the sentiment process.

Moreover, many of recent researches on SA in non-English languages are full machine translation –based (MT) to translate the whole tweets into English and not only used for lexicon expansion. These researches benefit from large English sentiment lexicons to find an appropriate polarity for any non-English word.

For example, [12] presented the multilingual sentiment analysis for twitter. They used an existing English sentiment system and adopted it to work with the Spanish language. First, they employed translation of English training data using Google translator into Spanish Language. Then, they extracted the unigram and bigrams features from translated (Spanish) data to train the SVM SMO classifier. They concluded that using linguistics processing (such as stop removal) is worsening the performance. In addition they obtained that using the four classifiers improved the performance of monolingual analysis.

In addition, there were a few recent studies that depended on fully machine translation approaches directed specially to Arabic language. For example, [13] presented a SA system based on translating the Arabic text to English. They used a Stanford sentiment classifier for English tweets (translated from Arabic tweets). They made a comparison between the proposed system and other lexicon and learning based sentiment systems. Their system provides a promising result with high accuracy of 76% compared to others. Similarly, [14] also approved by many experiments the efficiency of machine translation approaches. This approach provides a result similar to the state of art Arabic sentiment analysis system. They noticed that for any automatic translated text, the automatic sentiment Analysis is better than the manual sentiment analysis.

Finally, Table 1 provides you with the summaries and comparisons of twitter sentiments analysis related work based on different parameters. These parameters include study approach, language, polarity levels, sentiment features, data sets and the reported accuracy. In summary, the main challenges on twitter sentiment analysis are the domain and language dependency either on learning or lexicon based approaches. Regarding this, there were attempts to fix the language dependency using a machine translation approach by the Multilanguage system or translation for lexicon expansion. For the domain dependent challenge also there was an attempt to solve it by using the skipgrams terms without any language resources. We can see from the below table that the high accuracy achieved by using the semi-supervised approach with partial MT approach. After that hybrid lexicon based [7] and statistical learning based approach that has more features [2]. This means to enhance the performance of twitter sentiment analysis system; we need to build a system that is semi-dependency on language and domain.

Table 1. Comparison of related work.

Research	Approach	Levels	Sentiment features	Language	Lexicons/Data sets	Accuracy
[5]	• Unsupervised • **Lexicon-based**	4 + 6	• Opinion words • Hashatags • Emoticons • Negation	Spanish	Opinion words (2509 p, 5626 n) Hashatage 172p,127n Emoticons (70 p, 46 n) (32p, 38 p+, 23 n 23 n+)	• 61% • 51% with negation
[8]	• Unsupervised • **Lexicon-based** • Domain dependent	3	• Opinion words • Negation	Chinese	• Energy (1517) • Soccer (119) • Life (43) • World cup (34)	72.90%
[7]	• Unsupervised • **Lexicon-based** • Domain independent • Hybrid system	3	• Emoticons • SentiWord Net analysis • Opinion words	English	Emotions (145: 70 P,75 n) Opinion words (9493) Tweets (2116) randomly.	85.7%
[2]	• Supervised • **Learning-based** SVM, ME • Domain independent • Statistical	3	• Word ngrams • Character ngrams • All-caps: # • POS: # • Hashtags: # • Punctuation: # • Negation: # • Elongated words: # • Clusters • Emoticons: • Automatic lexicon features • Manual lexicon features	English	• SemEval-2013 datasets, • NRC Emoticon lexicon(2312p,3324n) • HT(19121p,20192n) • S140(	85.5%
[6]	• Supervised • **Leaning-based** SVM	3	• Skipgrams	Spanish	Tweets: • Train(5899), • Dev train (997), • Dev test(3813)	P: 0.636 R:0.625 F1:0.626

[4]	• Supervised • **Leaning-based** MXE and SVM	3	• N-grams • Character n grams • POS • Emotions • Variant TFIDF features	Czech	10,000 posts from facebook and also more from product and movie reviews	F1:0.69
[9]	• **Leaning-based** Naive bays • Election domain	3	• Opinion words	Indian	Election related downloaded tweets.	NA
[10]	• Semi-supervised • **Partial MT approach**	3	*Linguistic* • egation Is_Negation,N_O_Negation • Intensifiers • Questions Is_Question,N_O_Question • Supplication and Wishful *Syntactic* features for conflicting phrases POS tag, Bi-gram	Arabic and Egyptian delicate	2000 topics • 997 tweets 49%, • 300 hotel reservation comments 16%, • 346 product reviews 17%, • 357 TV-Programs comments 18%	95%.
[11]	• Un-supervised • Lexicon-based • **Partial MT approach**	3	• Emotions • Negation	Arabic	• 4400 tweets • Lexicon: 2376: (1777 N, 600 P).	Exp1:23% Exp2:46%
[12]	• Supervised • Learning-based SVM SMO • **MT approach** (Google Trans.) • Multilingual sentiment analysis	5 + 3	English • Replacement with polarity: Opinion words,Slang,Emoticons • Repeated punctuation ("?","!",".") • Modifiers; Negation,Intensifier,Diminisher • Classifier Unigram,Bigram	English (labelled data) to Spanish	• English tweets (15196) • Spanish tweets (95,863)	F1: 0.69 without Slang replacement

[13]	● Supervised ● **MT approach** (Google and Microsoft trans.)	2	Stanford Sentiment Classifier	Arabic to English	● Arabic tweets: 937 (470p, 467n)	76%
[14]	● **MT approach**	3	[2] system features	Arabic to English	Arabic tweets: Unique tweet 163,944 p 37,848 n	NA
[15]	● Deep learning ● **Lexicon -based**	3	N-grams, word embeddings, sentiment lexicons	English	80,000 tweets about COVID-19 218,791 tweets about vaccination	reach 93%
[16]	Attention-based LSTM	3	fastText word embedding	indonesian	38071 data lines in the electronic domain	83% and 74%

3. Arabic Twitter Sentiment Analysis

This Sentiment analysis of Arabic text in twitter has many difficulties and challenges. It is mainly affected by the nature of Arabic language as any Arabic text analysis. In addition to the challenges related to the properties of Twitter text. These problems related to Arabic language or Twitter which should be considered at the sentiment analysis are illustrated on the following points:

3.1. Challenges Related to Arabic Language:

Arabic is a more morphologically complex language than English. So, it is a highly inflectional and derivational language. This means that for a given root can take different forms depending on the context [17] as in example 1 of Table 2. Also each root can generate many words with different meanings as in example 2 of Table 3. Therefore, there are issues related to the lexical features that might be helpful in sentiment analysis. This provides some importance to using the stemming for Arabic sentiment analysis [18].

Table 2. Example of Morphological Analyser of Arabic.

	Root: حب	
(1)	Different forms	أحب ويحب ويحبون
(2)	Different meaning	حبوب وحبيب وحوبة

Arabic has various diacritics; so for any words with the same spelling, we can have different meanings depending on its diacritics. For example: (يَعْلَم) which means teaching and (يَعْلَمُ) punctuation such as which means know [19]. Twitter text mostly lacks such diacritics, so this will increase the unambiguous meaning of single words.

Unavailability of Arabic Sentiment Lexicons [20]. This is considered as a big challenge encountered by any researcher on Arabic sentiment analysis compared to other languages. This might be due to the limitation of Arabic review websites that could be helped to build the sentiment lexicons [18]. Therefore, the research should be directed to build a big and effective lexicon and corpora. On the other hand, we need to encourage participation in increasing the kind of Arabic review websites that will increase the Arabic content on the web in general.

3.2. Challenges Related to Twitter

- Most people on twitter use informal language. They usually write about their diary or reply to others. This slang language is updated frequently and differs between Arabic countries, so it is difficult to build stable opinion lexicons.
- Moreover, people sometimes write their own abbreviation as in example 3 of Table 3. Also having a lot of spelling errors or using words from many languages. They could also be writing English words in Arabic letters as in example 4 of Table 3 [19].The unstructured format of tweet text leads to difficulty in parsing and therefore understanding the meaning.
- Twitter text Includes special symbols for hashtages '#'and replaying to users '@'
- The twitter text is very short. It is about 140 letters only.

Table 3. Example Arabic abbreviation and English word in Arabic letter.

	Meaning in English	Arabic word
(3) Arabic abbreviation	I'll be right back	برب
(4) English word in Arabic letter	Valentine day	فالنتاين

4. Proposed Method

This paper contribution is a supervised learning-based approach that uses the words or terms of tweets as features without external resources. In addition, the system aims in future to use lexicon resources to extract more important features related to Arabic languages. In summary, our approach performs the preprocessing and stemming of tweets. Then, the system generates the features space of skip-grams terms from the whole dataset. After that, it represents each tweet by these skip-grams features and their weights. Finally, we will train the classifier to build the sentiment classification model. The Figure below illustrates the architecture of the proposed sentiment analysis system of Arabic twitter.
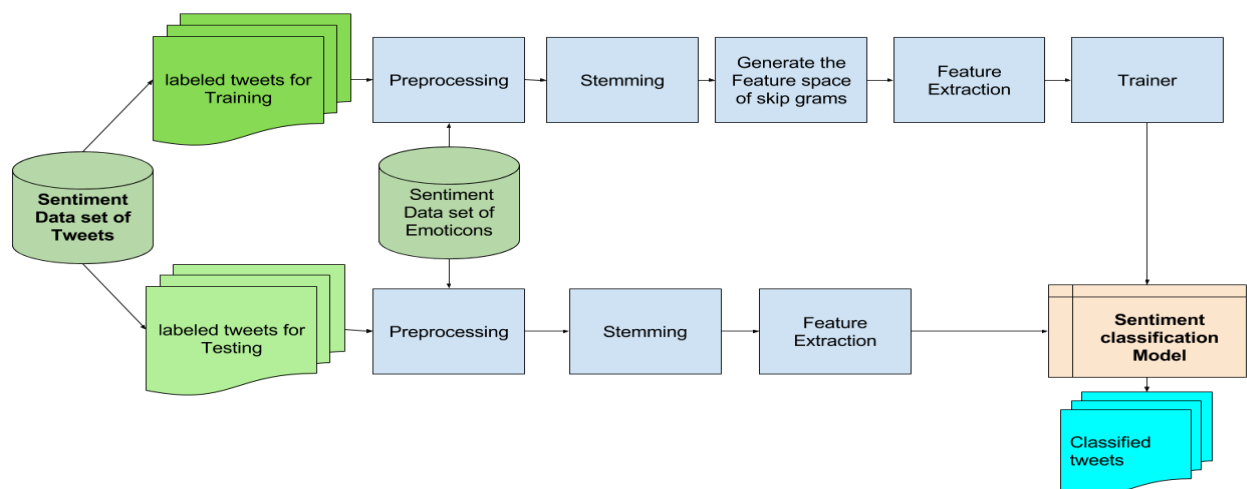
Figure 1. Proposed system Architecture of Arabic Twitter SA.

In the following subsections we will explain the main steps of our proposed method:
1. *Preprocessing:* This phase tends to prepare the tweets before generating the features of machine learning. It will include the following three steps:
a. Normalization:
Remove repetition: for each tweet, we remove the repetitive letters from a word if they are more than two words. For example, the word " جييييد" will become " جيد". This will help to find out the

correct stemming of the word. Moreover, the stemmed word would more easier to be found in the opinion lexicons if we could use it in future work.

Replacement with fixed tags: as mentioned above our proposed method is domain independent, so there is no need for a topic word, which is a hashtage in case of tweet. Therefore, we replace each hashtage by "HASHTAGE". In addition the username that marked by '@' is replaced by "USERNAME".

Replacement by polarity: we replace each emoticon found in tweet text by its polarity type ("NEGTIVE","POSITIVE" or "NETURAL"). This will be done using the emoticons sentiment lexicon as shown in the table 4 below:

Table 4. Emotions sentiment lexicon.

Positive Emoticons	Negative Emoticons			
%-)　(-:　(:	%-(　　)-:　　):　　)o:			
(^ ^) (^-^)	38*　　8-0　8/　　8\			
(^.^)	8c :#　　:'(　　:'-(			
(^_^) (o: *\o/*	:(　:*(　　:,(　　:-&			
--^--@	:-(:-(o)　:-/　　:-@			
0:)　8)	:-S:-\　　:-|　　:/			
:)　:　　:-)	:E :F　　:O　　:S			
:-* :-D	:[:[　　:\　　:_(　　:o(			
:-P　　:-} :3　:9	:s :|　　:Ö(　　</3　　<o< =(			
:D	=[　　>/			
:P :P :X :]	>:(>:L　>:O　　>[			
:b)　　　:o) :p	>\>o>　B(			
:fi	BcD:　　X(　　X-(			
;^)　　　<3	XO　　XP　　^o)　xP			
=) =]	|8C　　|8c			
>:)>:D				
>=D　　@}->				
-XD				
XP　　^_^				
x3?				
xD　　|D				
}:)				

b. *Tokenization:* this step extracts all the terms of the tweet without ignoring any kind of term.

c. *Stemming:* This is the most important part for Arabic: the complex morphologically language as we discussed in section 2. Here we use the Kojah stemmer [22] to find the stem for all the extracted terms of tweet. Table below shows an example of preprocessing steps.

2. Generate the feature space:　after having the stemmed tweet, we extract the skip-grams terms from each tweet by do the following:

a. Extract Skip-gram terms: the system obtains the skip-grams terms from each tweet. Skip-grams just like n-grams are formed but inw3 addition to allowing adjacent sequences of words, it also allows tokens to be skipped [6].

b. Calculate the skip-grams weights: for each skip-gram, the system calculates the weight depending on formula proposed by [6] as follow:

$$w(s,t) = \frac{terms(s)}{terms(s) + skips(s,t)}$$

where w(s,t) represents the weight of the skip-gram s in the text t. terms is a function that returns the number of terms in skip-gram s, and skips is a function that returns the number of skip-gram s in the text t.

c.   Generate features: each extracted skip-gram with its corresponding weight will be considered as a feature for machine learning.

3. *Extracting the feature:* as we have seen the output of the previous step is a set of all skip-gram terms as features. While in this step, the system represents new instance for each tweet based on the feature space or vector space model (VSM). Each tweet text represent as a vector, with values of term weights for each Skip-gram. Thus, given a set of N tweets t1, t2... tn, the tweet per term table is constructed as shown in table 5. In this table si is a skip-gram term and wij is the weight of skip-gram term j in tweet i. As the vectors for each tweet would have a very high dimension and would be very sparse, we create the sparse instance instead to reduce the dimensions and save the memory. The weight for the skip-gram terms was defined as on Table 5:

Table 5. Tweet by feature - Vector Table.

	s_1	s_2	....	s_i	n	Class
t_1	W_{11}	W_{12}	....	W_{1i}	W_{1n}	Positive
t_2	W_{21}	W_{22}	....	W_{2i}	W_{2n}	Negative
:						Neutral
t_i	W_{i1}	W_{i2}	....	W_{ii}	W_{in}	Positive
t_n	W_{n1}	W_{n2}	.	W_{ni}	W_{n}	Negative

4. *Trainer:* This step aims to build the sentiment classification model. We used the support vector machine SVM that has been considered as a best machine learning technique for text categorization.

5. *Classification:* here we used the classification model that has been built in the previous step to classify the new tweets. Each tweet labels to one of three levels: Positive, Negative or Neutral.

Table 5. Example of create skip grams from one tweet in dataset1.

Raw Tweet	@@LAT@@ شاهي وجلسة خطيرة تب كابي قصر السلطان #تركيا @@LAT@@ @@LAT@@ @@LAT@@)
Preprocess	[USERNAME, HASHTAGE, سلط, قصر, كبي, توب, خطر, جلس, شهي, USERNAME, USERNAME, USERNAME, ترك]
Skip grams	[USERNAME, USERNAME شهي, جلس, شهي جلس, شهي, شهي خطر, جلس خطر, جلس توب, جلس توب, خطر, USERNAME خطر توب, خطر كبي, توب, توب كبي, توب قصر, توب قصر, كبي, كبي قصر, كبي سلط, قصر, قصر سلط, سلط, HASHTAGE USERNAME ، HASHTAGE, سلط, سلط قصر HASHTAGE , HASHTAG, سلط , USERNAME, USERNAME USERNAME, HASHTAGE USERNAME, USERNAME]

```
          Building the classification model of Arabic twitter
                      sentiment analysis system


   Input: Annotated sentiment tweet corpus


   Output: sentiment classification model


        For each tweets in the data set


1. Reading the tweet text and their
polarity from dataset
2. Preprocessing
   2.1 Normalization: remove repetition,
replace hashtage and usernames with fixed
tag, replace emoticons with their
polarity.
   2.2 Tokenization
   2.3 Stemming: by Khoja stemmer
3. Generating the features: SkipGrams
   3.1 extract Skipgram terms
   3.2 calculate the skipgrams weights
   3.3 Generate features: weka attributes
      For each tweets in the data set


Creating the instances for all tweets
based on the feature space
   1. Extract Skipgram terms from current
tweet
   2. Get the index of each skipgram from
the feature space
   3. Set the weight at the index:
position of each skipgram attribute
 Build    the    classifier   based   on   new
instances: SVM
```
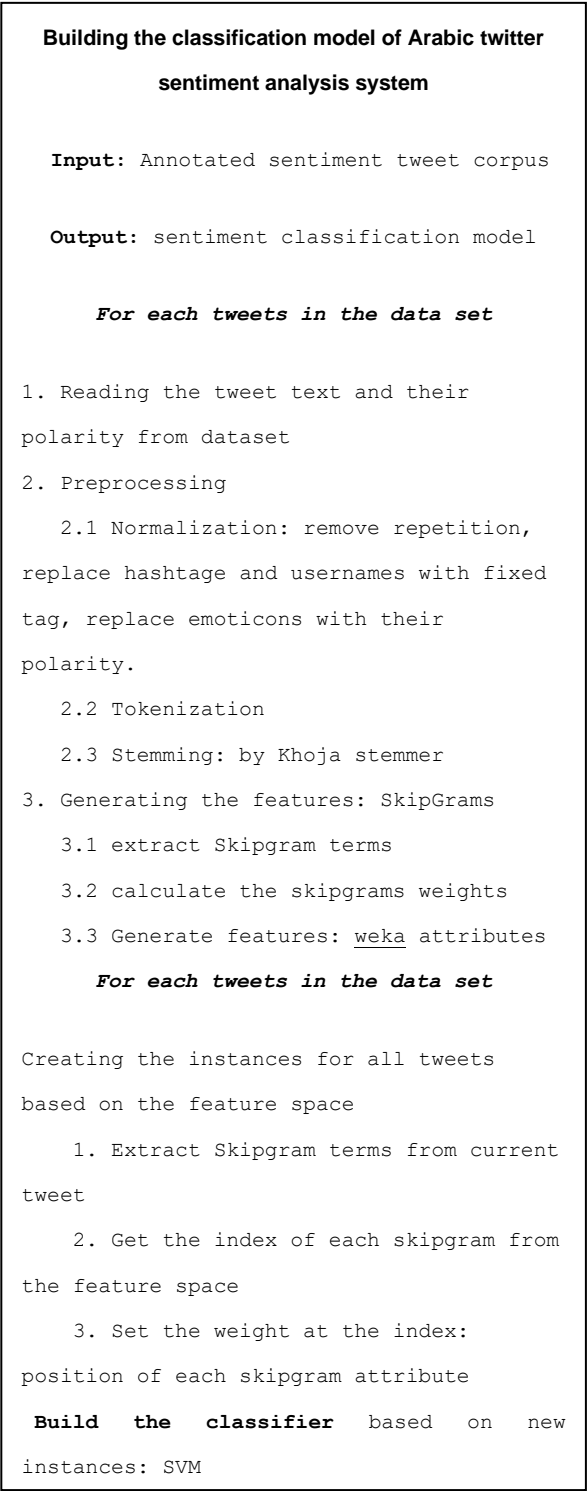
Figure 2. Algorithm of Proposed system.

5. Experiments and Results

To evaluate the ability of the proposed method to measure the sentiment analysis of Arabic tweets, we perform many experiments on two different datasets. The first data set is the Golden twitter sentiment corpus by Refaee and Rieser in 2014 [23]. The second dataset is the Arabic sentiment dataset ASTD by Nabil, Aly, and Atiya 2015 [24]. The size and distribution of polarities of these datasets is shown on Tables below.

Table 7. Dataset1 distribution in number of tweets.

Polarity	Number of tweets
Positive	876
Negative	1941
Natural	3697
Total	6514

Table 8. Dataset2 distribution in number of tweets.

Polarity	Number of tweets
Positive	799
Negative	1,684
Mixed	832
Objective	6,691
Total	10,006

We perform many experiments to build the sentiment classification model using the previous datasets. First we split our data set to train and test set. The train test ratio on both datasets is 66% for training and the remaining for validation and testing. All the experiments were implemented using java language to build our system with the help of weka API for machine learning. Also the experiments worked on a machine with CPU of Intel Core i7, 3.1 GHz speed and 16GB of memory.

To show the impact of stemming on sentiment analysis of Arabic twitter, we build two scenarios of experiments. The first scenario applies the stemmer while the second one without applying the stemmer. For the two scenarios we employed a naïve Bayes classifier in addition to five different support vector machine classifiers. The SVM classifiers are LIBSVM and SMO with different kernels types. LIBSVM is library for support vector classification and regression algorithm. This library was generated by chang and Lin in 2001 [25]. We used it in our experiment with the default parameters. Also we used another classifier of support vector machine, which is SMO. SMO implements the sequential minimal-optimization algorithm for training a support vector classifier, using kernel functions such as polynomial or Gaussian kernels (PBF) [25]. We used the SMO here with four different kernels: PolyKernel, NormalizedPolyKernel, RBFKernel and PUK. The results of the experiments are shown in the tables below. In these tables we present the weighted precision, weighted recall, weighted F-measure and the accuracy. In detail, we will discuss the experiments on the following subsections based on the dataset used.

Table 9. Numbers of features on each different experiment.

Datasets	Numbers of skip grams features
Dataset 1-stem	84802
Dataset 1-no stem	117436
Dataset 2-stem	109702
Dataset 2-no stem	

5.1. Experiments on Dataset 1:

Table 10 shows the results for six different classifiers that employ the first dataset based on two scenarios: stemming and not stemming. In the first scenario, the SMO classifier with polynomial kernels (PolyKernel) significantly outperforms other classifiers with highest Recall, F-measure and Accuracy (0.66,0.66 and 66.24). While the SMO with RBFKernel yields the best results based on Precision measure (0.68).

In the second scenario, the stemmer was not applying. Similarly with the first scenario, the best results are obtained by applying the PolyKernel of SMO with highest Recall, F-measure and Accuracy (0.64,0.63 and 64.20). In addition the NormalizedPolyKernel produces similar results of the polynomial kernel but with a smaller value of F-measure by two points only (0.64,0.61 and 64.20). In

a different way, the SMO with PUKKernel provides the highest Precision (0.67) instead of RBFKernel in the stemming case. In contrast, the naïve Bayes and SMO with NormalizedPolyKernel perform worse when there is no stemming. Surprisingly, the LIBSVM and SMO- PUKKernel provide the same smallest values over all performance measures in both cases.

    Generally, the experiments on this data set obtain that using the stemmer provides performance that outweighs the second case without the stemmer.

Table 10. Dataset 1 results.

Stemming	Classifier	Precision	Recall	F-measure	Accuracy
Yes	LibSVM	0.32	0.56	0.41	56.40
	Naive Bayes	0.63	0.62	0.62	62.21
	SMO-PolyKernel	0.66	**0.66**	**0.66**	**66.24**
	SMO-NormalizedPolyKernel	0.67	0.65	0.61	65.38
	SMO-RBFKernel	**0.68**	0.64	0.58	63.93
	SMO-PUK	0.67	0.57	0.42	56.90
No	LibSVM	0.32	0.56	0.41	56.40
	NaiveBays	0.59	0.60	0.59	60.03
	SMO-PolyKernel	0.63	**0.64**	**0.63**	**64.20**
	SMO-NormalizedPolyKernel	0.65	**0.64**	0.61	**64.20**
	SMO-RBFKernel	0.66	0.61	0.52	60.89
	SMO-PUK	**0.67**	0.57	0.42	56.85

5.2. Experiments on Dataset 2:

    In these experiments we used the dataset2 for the two scenarios as shown on Table 11.

    In the first scenario, the SMO classifier with NormalizedPolyKernel significantly outperforms other classifiers with highest Precision, Recall and Accuracy (0.77,0.76 and 76.23). While the SMO with PolyKernel yields the best results based on F- measure (0.71).

    In the second scenario, the stemmer was not applying. In contrast with the first scenario, the best results are obtained by applying the PUK of SMO with highest Precision, Recall and Accuracy (0.74, 0.76 and 76.04). Similarly with the first case, PolyKernel achieves the highest F-measure with 0.69. Interestingly, the LibSVM, NormalizedPolyKernel and RBFKernel provide similar highest recall with PUK (0.76). In the same way, LIBSVM remains on the same performance for the two scenarios. While the naïve Bayes shows better performance for not stemming case rather than the stemming case.

    At the end of these experiments we can conclude that stemming improve the result of Arabic Twitter text classification. In addition, the SMO is better than LIBSVM. The best kernals used are PolyKernel then PUK, NormalizedPolyKernel and finally RBFKernel. We can also observe that the use of a large data set as in dataset2 will provide us with more higher accuracy (76.23) than dataset1 (66.24). Finally, our result is better than [6] that used the same skipgrams terms on sentiment analysis of twitter but on Spanish language. This might be due to our using of the emoticons replacement with its polarity. There we have to provide more importance of replacement with polarity method but for the opinion words. It can be applied by finding the external resource of Arabic opining lexicons.

Table 11. Dataset 2 results.

Stemming	Classifier	Precision	Recall	F-measure	Accuracy
Yes	**LibSVM**	0.58	**0.76**	0.66	75.99
	Naive Bayes	0.71	0.58	0.62	58.43
	SMO-PolyKernel	0.69	0.73	**0.71**	73.06
	SMO-NormalizedPolyKernel	**0.77**	**0.76**	0.66	**76.23**
	SMO-RBFKernel	0.58	**0.76**	0.66	75.96

	SMO-PUK	0.74	**0.76**	0.66	76.05
No	**LibSVM**	0.58	**0.76**	0.66	75.98
	NaiveBays	0.68	0.67	0.67	66.93
	SMO-PolyKernel	0.67	0.72	**0.69**	71.97
	SMO-NormalizedPolyKernel	0.67	**0.76**	0.66	76.01
	SMO-RBFKernel	0.58	**0.76**	0.66	75.98
	SMO-PUK	**0.74**	**0.76**	0.66	**76.04**

6. Conclusions

In this paper we presented a review of Twitter sentiment research. These studies can be divided into three categories based on their approach: Lexicon-based, learning-based and Machine translation–based (MT). The main challenges on twitter sentiment analysis are the domain and language dependency either on learning or lexicon based approaches. Therefore, there were attempts to fix the language dependency using a machine translation approach by Multilanguage system or translation for lexicon expansion.   In fact, Sentiment analysis of Arabic text on twitter has many more challenges. It is mainly affected by the nature of Arabic language and the challenges related to the properties of Twitter text. These problems related to Arabic language or Twitter should be considered at the Arabic sentiment analysis.

The proposed method is a supervised learning-based approach that uses the words or terms of tweets as features without external resources. This system with Skipgrams method achieves up to 76.23% accuracy that is better than previous work using the same techniques. Our experiments show some important results. First, stemming has a big impact for Arabic twitter sentiment classification. Secondly, the best learning-technique is the SMO with polyKernal. Finally, the replacement with polarity method is more effective to use as we observed from Emoticon replacement.   Under the lack of Arabic sentiment resources, this work is promising to have improvement and solve its limitations. In future, the system aims to use lexicon resources to extract more important features related to Arabic languages such as replacement of opinion words with its polarity.

References

1.   B. Liu, 'Sentiment analysis and subjectivity', in Handbook of Natural Language Processing, Second Edition. Taylor and Francis Group, Boca, 2010.
2.   S. Kiritchenko, X. Zhu, and S. M. Mohammad, 'Sentiment Analysis of Short Informal Texts', J Artif Int Res, vol. 50, no. 1, pp. 723–762, May 2014.
3.   L. Williams, C. Bannister, M. Arribas-Ayllon, A. Preece, and I. Spasić, 'The role of idioms in sentiment analysis', Expert Syst. Appl., vol. 42, no. 21, pp. 7375–7385, Nov. 2015.
4.   I. Habernal, T. Ptáček, and J. Steinberger, 'Reprint of "Supervised sentiment analysis in Czech social media"', Inf. Process. Manag., vol. 51, no. 4, pp. 532–546, Jul. 2015.
5.   S. M. Jiménez Zafra, E. Martínez Cámara, M. T. Martín Valdivia, and L. A. Ureña López, 'SINAI-ESMA: An unsupervised approach for Sentiment Analysis in Twitter (PDF Download Available)', presented at the TASS, 2014.
6.   J. Fernández, Y. Gutiérrez, J. M. Gómez, and P. Martinez-Barco, 'GPLSI: Supervised Sentiment Analysis in Twitter using Skipgrams', in Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), Dublin, Ireland, 2014, pp. 294–299.
7.   F. H. Khan, S. Bashir, and U. Qamar, 'TOM: Twitter opinion mining framework using hybrid classification scheme', Decis. Support Syst., vol. 57, pp. 245–257, Jan. 2014.
8.   L. Lin, J. Li, R. Zhang, W. Yu, and C. Sun, 'Opinion Mining and Sentiment Analysis in Social Networks: A Retweeting Structure-Aware Approach', in 2014 IEEE/ACM 7th International Conference on Utility and Cloud Computing (UCC), 2014, pp. 890–895.
9.   G. P.Wani and N. V. Alone, 'Analysis of Indian Election using Twitter', Int. J. Comput. Appl., vol. 121, no. 22, pp. 37–41, Jul. 2015.

10. H. S. Ibrahim, S. M. Abdou, and M. Gheith, 'Sentiment Analysis For Modern Standard Arabic And Colloquial', Int. J. Nat. Lang. Comput., vol. 4, no. 2, pp. 95–109, Apr. 2015.

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.

12. A. Balahur and J. M. Perea-Ortega, 'Sentiment analysis system adaptation for multilingual processing: The case of tweets', Inf. Process. Manag., vol. 51, no. 4, pp. 547–556, Jul. 2015.

13. E. Refaee and V. Rieser, 'Benchmarking Machine Translated Sentiment Analysis for Arabic Tweets', NAACL-HLT 2015 Stud. Res. Workshop SRW, p. 71, 2015.

14. M. Salameh, S. M. Mohammad, and S. Kiritchenko, 'Sentiment after Translation: A Case-Study on Arabic Social Media Posts', in North American Chapter of the Association for Computational Linguistics (NAACL), 2015.

15. Ainapure, B.S.; Pise, R.N.; Reddy, P.; Appasani, B.; Srinivasulu, A.; Khan, M.S.; Bizon, N. Sentiment Analysis of COVID-19 Tweets Using Deep Learning and Lexicon-Based Approaches. Sustainability 2023, 15, 2573. https://doi.org/10.3390/su15032573

16. Ismet, Hilya Tsaniya, Tanzilal Mustaqim, and Diana Purwitasari. "Aspect based sentiment analysis of product review using memory network." Sci. J. Informatics 9.1 (2022): 73-83.

17. R. M. Duwairi, R. Marji, N. Sha'ban, and S. Rushaidat, 'Sentiment Analysis in Arabic tweets', in 2014 5th International Conference on Information and Communication Systems (ICICS), 2014, pp. 1–6.

18. L. Albraheem and H. S. Al-Khalifa, 'Exploring the Problems of Sentiment Analysis in Informal Arabic', in Proceedings of the 14th International Conference on Information Integration and Web-based Applications & Services, New York, NY, USA, 2012, pp. 415–418.

19. S. O. Alhumoud, M. I. Altuwaijri, T. M. Albuhairi, and W. M. Alohaideb, 'Survey on Arabic Sentiment Analysis in Twitter'.

20. S. R. El-Beltagy and A. Ali, 'Open issues in the sentiment analysis of Arabic social media: A case study', in 2013 9th International Conference on Innovations in Information Technology (IIT), 2013, pp. 215–220.

21. E. Refaee and V. Rieser, 'Subjectivity and sentiment analysis of arabic twitter feeds with limited resources', Workshop Free.-Source Arab. Corpora Corpora Process. Tools Workshop Programme, p. 16, 2014.

22. S. Khoja and R. Garside, 'Stemming arabic text', Lanc. UK Comput. Dep. Lanc. Univ., 1999.

23. E. Refaee and V. Rieser, 'An Arabic twitter corpus for subjectivity and sentiment analysis', Proc. Ninth Int. Conf. Lang. Resour. Eval. LREC'14 Reyk. Icel. May Eur. Lang. Resour. Assoc. ELRA, 2014.

24. M. Nabil, M. Aly, and A. Atiya, 'ASTD: Arabic Sentiment Tweets Dataset.', 2015, pp. 2515–2519.

25. I. H. Witten and E. Frank, Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann, 2005.