Feel the Heat: Emotion Detection in Arabic Social Media Content

Omneya Rabie
Mentor Graphics Corporation
78 Elnozha St., Heliopolis, 11361 Cairo, Egypt
omneya_rabie@mentor.com

Christian Sturm
Hamm-Lippstadt University of Applied Sciences
Maker Allee 76-78, 59063 Hamm, Germany
Christian.sturm@hshl.de

ABSTRACT

The automatic detection of emotions in textual parts of social media websites such as Facebook and Twitter has applications for business development, user interface design, content creation, emergency response, among others. Current research has shown that it is possible to detect emotions for English content. To our knowledge, however, there are only few attempts for Arabic content. There is neither Arabic corpus with instances labeled for emotions, nor studies to detect emotions from Arabic microblogs content. Therefore, we collected Arabic text messages from the social networking website Twitter from January/February 2011. Human annotators labeled them with the corresponding emotions. Working with that corpus, our experiments show that emotions can be automatically detected from tweets after performing Arabic language related language preprocessing steps. Our contribution consists in adding preprocessing steps that have improved the classification results by 4.4% compared to the original Khoja stemmer. In addition, we have extracted a sample word-emotion lexicon from that corpus. Our experiment demonstrates that this sample word-emotion lexicon enhances the emotion detection results by 22.27% compared to the SMO classification using the train/test option. Finally, we show that the communication style used by the writer significantly relates with the emotion expressed in the text.

KEYWORDS

Emotion detection, Social media, Twitter, Arabic text, Classification.

1 Introduction

Long before awareness of the World Wide Web became wide-spread, people used to seek their friends' opinions or to consult consumer reports about products or services that they want to buy. Moreover, they would make verbal surveys in order to know which candidate most of the people are planning to vote for in a local election. Lately, the Internet and the Web have made it easier to collect such data. Text is widely used in the communication between people on the web. It delivers informative content, one's opinion, and emotional state. Microblogs allow a vast pool of people that are neither personal acquaintances nor well-known professional critics to provide their experiences, opinions, and emotions.

Studies about emotions have been conducted by psychologists and behavior scientists for long time [1] [2] [3] [4]. They consider emotions a major element of the human nature. Many sentiment analysis studies have been conducted to annotate text as positive/ negative valence. However, emotion annotation can be more effective and accurate. Despite the evolving importance and usage of microblogs, there only exist few trials for building microblog corpus for emotions labeling.

Although the increasing usage of Arabic language in blogs and social networks, it is not given enough focus in the sentiment analysis or emotion detection fields. According to [5], the number of Internet users in the Middle East was 90,000,000 in 2012. In addition, the Arabic language comes at the seventh place between the ten most used languages on the Internet. It is expected to take the fifth place in 2015. The Arabic content is in continuous increase in social media.

We are contributing to the research area of emotion detection in Arabic text by describing the experiments concerned with emotion analysis of Arabic microblogs content. The main idea is to automatically detect the emotions of microbloggers who use Arabic language. For this purpose, an Arabic corpus was built out of microblog text. Different classification techniques were evaluated. Moreover, a set of features was checked through our experiments. The aim of this project is to highlight the importance of the Arabic social media content analysis and going beyond polarity classification of text, to conduct automatic emotion analysis based on the most basic emotions suggested in [6] and [2].

This paper is structured as follows: section **2** includes some related work in emotions and sentiment analysis fields. Then, section **3** describes our experimental setup and methods for analyzing Arabic social media content gathered from Twitter. Section **4** contains the results of the different classification algorithms and the comparison between several preprocessing techniques. Finally, we conclude with a discussion of the implications that this work has for the automatic emotion detection task.

2 Related Work

"Sentiment Analysis is the computational treatment of opinion and subjectivity in text." [7]. It is the application of natural language processing and text analytics to determine the target information from the text. It is used to determine the attitude of the writer, including his/her evaluation of the topic as well as his/her emotional state with respect to some topic.

2.1 Sentiment Analysis of English Language

In the field of sentiment analysis, the work related to text polarity can be divided into two groups; **techniques to automatically generate sentiment lexicons, and systems that analyze sentiment in text documents**. The first group depends widely on adjectives. The lexicon construction starts with small seed lists and an assumption by the researcher. The hypothesis by Hatzivassiloglou and Mckeown in [8] was that if adjectives are separated by "and", they are of the same polarity; however if they are separated by "but", they are adjectives of opposite polarity. Later on, the gradation categorization was taken into consideration as well as polarity using statistical models [9]. WordNet lists [10] have been also used as seeds in the process of generation of positive and negative word lists. The assumption is that synonyms of a certain word have the same polarity, while antonyms have opposite polarity [11].

The second group is focused on **sentiment analysis systems**. The sentiment analysis of movie reviews in [12] was performed by a machine learning technique that achieved an accuracy of 83% for polarity classification. It was intended to consider subjective character to enhance sentiment analysis. However, in [11] the news sentiment analysis system considers facts and opinions as contributors to the public sentiment. The system focuses on local sentiments as more reliable than global document sentiment. The same approach was also adopted by [13].

The application in [11] concluded the names of people talking positively versus those who talk negatively about seven topics targeted in news and blogs. The classification was based on a list of words (lexicon) created during the experiment. The application in [14] defines the polarity of emotions in children fairy tales text using a linear classifier. Groups of annotators labeled the stories according to the six basic emotions of Ekman. Afterwards, all emotions were grouped under two classes: positive and negative emotions. The classification accuracy reached 69.37%. Different approaches of assigning positive/ negative score to a word out of context were compared in [15]. SVMs and Gaussian Processes were used to test the performance of all the metrics in conjunction. The study suggests a learning approach that combines the various formulae that compute prior posterior polarity. According to [16], for the more difficult multiclass case including a neutral class, accuracy is often below 60% for short messages on Twitter, the social media website. In the later study, the negation of different sentiments was taken into consideration. This practice improved the polarity classification by 5.4% compared to the methods that use bag of words features with NaïveBayes and SVMs.

In the field of emotion detection, the study in [17] considered the identification of the basic emotions in news headlines. Different systems were tried. The first system checks the presence of the WordNet Affect emotion word in the headline and categorizes the text accordingly. The second one was based on the annotation of blogs and the Bayesian classifier. The third system was based on the semantic of the words and allowed the detection of emotion related words. In [18] the author suggested including a semantic feature in the process of sentiment analysis of tweets that enhanced the accuracy of the classification by 6.5%.

The study in [19] targeted the analysis of the tweets containing an emotional hashtag based on the six basic emotions. An emotion labeled tweets corpus has been created (TEC) as well as a word-emotion lexicon. The results suggest that the TEC, after applying a domain adaptation technique, produced better results than the methods used in [17] even when applied in different domains. In addition, Twitter emotional hashtags were used by Gurini et al. [20] to design a recommendation system of friends that share the same sentiment about the topics of interest of each user.

2.2 Sentiment Analysis of Arabic Content

Some studies have been conducted in the field of sentiment analysis of Arabic content. The study in [21] focused on comparing the performance of different classifiers for the Arabic language datasets. Another study [22] compared the performance of the Sequential Minimal Optimization (SMO) and Naïve Bayesian (NB) in the classification task. These algorithms are the suggested classifiers based on excel in their performance; SMO comes first, and then NB. This excel was also highlighted in a polarity classification in [23]. Determining the polarity of opinions in Arabic documents was also the target of [24]. The study proposed a combination between three different classification methods; a lexicon based classifier, followed by the maximum entropy classifier, then K-nearest neighbor classifier. This approach achieved 80% performance accuracy. In addition, a classifier for social networks slang Arabic content was suggested in [25]. This study used 1350 comments as a dataset collected from news channels websites in order to use the SVM classifier to categorize tweets into positive/ negative classes. The training lexicon used was augmented by additional slang terms that enhanced the accuracy by 14%. Another study was concerned with the emotion detection based on the six basic emotions in the Arabic children stories using a computational approach [26]. This study mentions the importance of the word-emotion lexicon and the preprocessing steps that consider the punctuations and the negative words.

The Arabic approaches have some limitations and shortcomings, to our knowledge there is no developed Arabic corpus with instances labeled for emotions. The slang Egyptian dialect has not been investigated yet. Therefore, we took the approach described in the following section in order to cover some of these limitations.

3 Methodological Approach

The analysis of the data involved several steps: the dataset composition, the annotation process, the data preprocessing, and finally, different classification techniques were tried over the data.

3.1 Data Collection

We collected our data from Twitter, the social media website. The corpus collected includes 1776 tweets from more than 200 users, covering an 18 days period from January 25, 2011 to February 11, 2011. As the usages of words vary from a topic to another [19], we have chosen the Egyptian revolution in 2011 as the topic of concern. It is identified by the hashtag #jan25 on twitter. Tweets ranged from a one word tweet to 140 characters tweet.
We filtered out non Arabic tweets, retweets, tweets including photos or videos. Finally, the corpus was ready for the annotation process.

3.2 Data Annotation

We created surveys where the annotator task was to guess the emotion of the writer based on the provided tweet text. Three annotation runs took place. The 1st run included random Arabic tweets chosen from the collected set. It resulted in 1012 annotated tweets out of 1130 input tweets. The 2nd run included mainly some limited features tweets and it resulted in 609 annotated tweets out of 646 input tweets. The 3rd run was a confirmation run together with an annotation of the tweets according to the communication style used by the tweet writer; aggressive, assertive, or passive style.

The entire tweets collected from Twitter represent an Arabic emotion annotated tweets corpus. The annotators were Egyptians who witnessed the Egyptian revolution in 2011. An average of 15 persons labeled each tweet with the corresponding emotion. The emotions provided are the six most basic emotions [2] [6]. We have excluded the annotated tweets with less than 50% annotators' agreement. Finally, we constructed the Twitter corpus that consisted of 1605 tweets.

3.3 Data Preprocessing

The data preprocessing took place using five different techniques; basic preprocessing, basic preprocessing in addition to the removal of a list of stopwords, Lucene light Arabic stemmer [27], Shereen Khoja Arabic stemmer [28], and modified Khoja Arabic stemmer. The basic preprocessing includes the removal of non Arabic letters, multiple spaces, and punctuation. The list of stopwords is composed of a standard list of standard Arabic words [29]. We have also added to it their equivalents in the Egyptian slang dialect and some additional slang words that appeared in the collected dataset and have no emotion significance. The Lucene light Arabic stemmer eliminates the definite articles and few prefixes and suffixes only. The Khoja stemmer does all the previous functionalities in addition to reducing each word to its root; however, it handles the standard Arabic language only. We modified Khoja stemmer in order to include the Egyptian slang dialect.

3.3.1 Modified Khoja Stemmer

Preprocessing in Arabic language is of great use; especially, each Arabic word can be reduced to its root. The Khoja Stemmer is concerned with affixes removal as well as reducing words to their roots. It includes stemmer text files that contain an Arabic dictionary of roots. Also, lists of stopwords, affixes, as well as a list of strange words are included. After stemming, words should equate one of the roots provided by the stemmer. The Khoja stemmer is designed to support the standard Arabic language only. Our modifications aimed at customizing the Khoja stemmer to support the Egyptian slang dialect contained in our created corpus. Word segmentation has also been applied; the delimiter used is the space. Other symbols like (.), (,), (;), (:), multiple spaces, and new lines are replaced with one space. The Egyptian slang dialect has some common characteristics that can be emphasized. We handled them the following way:

- **Negation patterns**

In standard Arabic language negated words are preceded by a negation tool or word. For example, the verb hears "YASMA'A" in Arabic is negated if it is preceded by the word no "LA" to be "LA YASMA'A", doesn't hear. However, in slang negation it will be "MAYSMA'ASH". This pattern is formed by the addition of the letter ”MEEM” at the beginning of the verb and the letter ”SHEEN” at the end of the verb. This pattern has different forms for singular, plural and masculine, feminine person; it is shown in table **1**.

Table 1. Negation pattern.

	Subject	Pattern
Present tense	Singular masculine	مبيفعلش
Present tense	Singular feminine	مبتفعلش
Present tense	Plural masculine	مبيفعلوش
Present tense	Plural feminine	مبتفعلش or مبيفعلوش
Past tense	Singular masculine	مفعلش
Past tense	Singular feminine	مفعلتش
Past tense	Plural masculine	مفعلوش
Past tense	Plural feminine	مفعلتش or مفعلوش
Order	Singular masculine	متفعلش
Order	Singular feminine	متفعليش
Order	Plural masculine	متفعلوش
Order	Plural feminine	متفعلوش

We added to Khoja stemmer a Java method that checks the occurrence of such cases. If it is found, the word is reduced to its root instead of being ignored by the original Khoja stemmer. Moreover, it will be preceded by a negation word in the output file to differentiate between the occurrence of word and its negation.

- **Additional suffixes, prefixes, and stopwords.**

We added some slang suffixes and prefixes to Khoja files in order to remove them from the stemmed words through the reduction process. Stopwords are also removed from the text being stemmed. The list of suffixes, prefixes and stopwords is shown in tables **2**, **3**, and **4**.

Table 2. The list of additional suffixes.

Suffixes
لهم، لنا، نكم، ينك كي، اك ل، و، ي

Table 3. The list of additional prefixes.

Prefixes
ح، ه، حي، هي حت، هت حن، هن

Table 4. The list of additional stopwords.

Stopwords
ليا، لهن، لهم، له، لكي، لك
دول، ديه، دي، دا، دة، ده
دلوقتي، دلوقت، الآن، الان
انتي، باني
وكنتوا، كنتوا، وكنتو، كنتو، وكنتم، كنتم، وكانوا، كانوا، وكنت، كنت
بوك، فيس، تويتر، اللي، شوية، بس

- **Additional definite articles.**

For abbreviation, some letters are sometimes added before the definite article "AL". For example, the letter "AIN" to mean on top of, or the letter "SEEN" to mean the night of. Hence, a new form of the article appears. We added these forms to the definite articles text file in the stemmer.

- **Slang representation of periodontal letters.**

The Arabic language includes three periodontal letters, where two of them are replaced by other letters in the Egyptian slang language. The letter "THEH" is replaced by the letter "TEH", and the letter "THAL" is replaced by the letter "DAL". We added a Java method to check this case in order to reach the correct root of the word.

Moreover, three semantic properties were supported:

- **Word pairs**

In Arabic language some word pairs like "ALHAMD LELLAH" which means thanks God and "ALLAH AKBAR" which means God is the greatest, are used to express gratitude to God and happiness. Mistakenly, sometimes people write them as a single word. To avoid stemmer misinterpretation of such word pairs, we added a method to the code that checks the presence of these pairs and replaces them with same word in the output file.

- **Negation form**

As mentioned in the negation pattern property described earlier, any word preceded by a negation tool or is in the slang negation pattern form should not be replaced by its root. Instead, it should be replaced by the negation of this word in the output file. We added a negation words list (standard and slang), so that it could be checked during the reduction process. If the algorithm finds a word of this list in front of another word, both will be replaced by the root of the word preceded by a specific negation word.

Table 5. The list of negation words that are replaced with the negation word (لا) in the output file:

Negation words
بلا، فلا، ولا، لا
فلم، ولم، لم
فلن، ولن، لن
فليس، وليس، ليس
منغير، بغير، بدون

- **Disapprobation words**

In the Egyptian slang language, some words are used to express the same meaning. For example, "HOWA EZAY", "HOWA EIH" both denotes disapprobation. The form of these words is usually a word denoting a question followed or preceded by a stopword, or a negation word. We added a method in the code to check for such words and replace them with the same word in the output file.

These additions were adopted in order to facilitate the classification job and enhance its performance. Using the modified Khoja stemmer as the preprocessing technique for the first annotation run tweets (1012 tweets), we extracted some emotion related attributes in order to form a seed list for a sample word-emotion lexicon.

3.4 Sample Word-Emotion Lexicon

Weka selects attributes option was used in order to extract the most effective features included in the first 1012 annotated tweets. The feature selection algorithm used was **BestFirst**. The extracted features formed the base of a word-emotion lexicon. In addition, the lexicon has been extended with some manually crafted emotion related words that were manually extracted from these tweets. The direct annotation of words performed in this way usually performs better than other methods according to [19]. The created lexicon was used in the collection of the second annotation run tweets. This sample lexicon was used in order to create a limited features environment for further experiments as will be shown in section **4**.

3.5 Data Classification

Finally, different data classification techniques were tried out. Weka software has been used for the SMO classifier, which is a simplification of the SVM classifier, and for the NaïveBayes classifier. And, a simple search and frequency algorithm based on the extracted sample word-emotion lexicon was also created. This algorithm counts the number of each emotion related words in the tweet. Then, it decides the emotion category of the tweet based on the emotion receiving the highest count. Ten folds cross validation was applied for the learning based algorithms. A comparison between the performances of the different classification algorithms has been held. Further, we conducted a comparison between the effects of the five different preprocessing techniques. Two different environments have been tried the random tweets environment and the limited features tweets environment.

4 Experiments and Results

The evaluation of the performance involves the calculation of the precision (P) and recall (R). They are calculated as follows:

P = #correct guesses / #total guesses (1)

R = #correct guesses / #total (2)

The number of correct guesses is the number of tweets marked correctly as expressing an emotion X by the classifier. The total guesses, is the total number of tweets that are marked by the classifier as expressing the emotion X (including correct and wrong guesses). The total number is the number of tweets expressing the emotion X in the dataset. Moreover, F is the balanced F-score which is calculated using the following formula **(3)**.

F = 2PR / (P+R) (3)

4.1 Random Tweets Tests

The first 1012 tweets that were randomly chosen were subject of different tests. They consist of 259 anger tweets, 127 disgust tweets, 149 fear tweets, 271 happiness tweets, 110 sadness tweets, and 96 surprise tweets. The first test was the comparison between the classification performance of the SMO and NaïveBayes classifiers.

4.1.1 SMO vs NaïveBayes

The first 1012 tweets were stemmed using the modified Khoja stemmer. Cross validation test with 10 folds run over these tweets using both classifiers. The weighted average results of both classification algorithms are shown in table **6**. Comparing the overall performance of both classification algorithms, the experiment demonstrates that the SMO classifier outperforms the NaïveBayes classifier by almost 5%. Therefore, in the rest of the tests, the SMO classifier was chosen.

Table 6. The weighted average results of SMO vs. NaïveBayes performance.

Classifier	P	R	F
NaïveBayes	0.399	0.391	0.394
SMO	0.442	0.451	0.441

4.1.2 The Five Preprocessing Techniques

The five preprocessing techniques have been compared. The weighted average classification results using the basic preprocessing (BP), basic preprocessing and removal of stopwords (BPRS), Lucene light Arabic stemmer (LLAS), Khoja Arabic stemmer (KAS), and modified Khoja Arabic stemmer (MKAS) are represented in table **7**.

Table 7. The weighted average classification results of the first 1012 random tweets with different preprocessing techniques.

Preprocessing Technique	#attributes	P	R	F
BP	6007	0.423	0.418	0.402
BPRS	5875	0.41	0.409	0.387
LLAS	5371	0.423	0.418	0.402
KAS	1283	0.394	0.409	0.397
MKAS	1450	0.442	0.451	0.441

BP technique resulted in 6007 attributes. The classifier took advantage of the variety of attributes and related a large number of features to each emotion category. That was done even with the words that do not really have emotion significance, for example, the words contained in the stopwords list. BPRS technique resulted in 5875 attributes. The attributes that were erroneously considered related to emotions in BP technique, were unique for each corresponding emotion class. When they were removed in BPRS technique, fewer words were associated to each emotion class, hence, a decrease in the overall performance of the classifier.

LLAS technique only removes the definite articles and few prefixes and suffixes. Therefore, the number of attributes decreased to 5371. However, this reduction did not enhance the overall classification results compared to the results of BP technique. The KAS technique reduced the number of attributes to 1283. This reduction is due to the stopwords list eliminated, in addition to the reduction of each word to its root. In this manner, after stemming, many words are represented with the same root, hence, the same attribute. Moreover, this technique removes any non stemmed word even if it is not from the stopwords list except the list of the strange words, which has been defined in the stemmer files before the test. Therefore, the classifier did not depend on relating many attributes to the emotion category. Instead, it depended on less attributes that were more frequently repeated in the different tweets. The MKAS that we developed takes into consideration the Egyptian slang language words that occurred in the tweets as mentioned in section **3**. Thus, the number of attributes that resulted from the implementation of this technique is 1450 attributes. This increase, compared to KAS technique can be related to the inclusion of some Arabic slang words that were removed by the KAS technique. Compared to BP technique, our MKAS technique enhanced the classification results by 3.9%. While, compared to the original KAS, it resulted in 4.4% enhancement in the overall classification results.

The attributes specified in the preprocessing stage are the features that the classification mainly depends on. We have chosen from them the most significant attributes as mentioned in **3** to be the seeds of a sample Arabic word-emotion lexicon.

4.2 Limited Features Tweets Tests

The random choice of tweets compared to the size of the sample set collected might not show the classification performance neatly. Therefore, the tweets that include the extracted features (the sample word-emotion lexicon) were selected to form a set of limited features tweets. The assumption is that allowing more repetition of words and separability of features related to each class would compensate for a larger dataset. Thus, the concept can be generalized for bigger size sets. Tweets containing words from the sample word-emotion lexicon were grouped together. A total of 1000 limited features tweets collected from the total 1605 tweets set that were finally available. They have been subject of the same tests as the previously selected random tweets.

4.2.1 The five preprocessing techniques

We checked the effect of the different preprocessing techniques on the classification performance for the limited features tweets as shown in table **8**. Taking the BP technique results as a reference, the usage BPRS technique enhanced the results by 1.1%. LLAS technique resulted in 1.7% enhancement. KAS technique increased the overall performance by 14.6%. Moreover, it has been shown that the MKAS enhanced the results by 17% compared to the classification using BP technique. Thus, our modified Khoja stemmer still outperforms the other preprocessing methods.

Table 8. The weighted average classification results of the limited features tweets using the five different preprocessing techniques.

Preprocessing Technique	Attributes	P	R	F
BP	5952	0.555	0.553	0.539
BPRS	5833	0.571	0.562	0.55
LLAS	5331	0.568	0.566	0.556
KAS	1277	0.691	0.685	0.685
MKAS	1433	0.716	0.71	0.709

Finally, the total dataset was classified using the SMO classifier. Moreover, we checked the performance of the simple search and frequency algorithm based on the sample word-emotion lexicon.

4.3 Total tweets set test

The total tweets set consists of 409 anger tweets, 204 disgust tweets, 285 fear tweets, 340 happiness tweets, 201 sadness tweets, and 166 surprise tweets. It was classified using the SMO classifier with cross validation option. It resulted in a weighted average balanced f-measure of 0.531, weighted average precision of 0.535, and weighted average recall of 0.535.

The SMO classifier with train/test option was investigated. The 1012 tweets of the first run were given to the algorithm as training set. In addition, the second run tweets (609 tweets) were entered as testing data. The MKAS preprocessing technique was used. The results of using the sample word-emotion lexicon search and frequency (SF) algorithm in the classification of the same 609 tweets were compared to the SMO classifier results. As shown in table **9** the experiment demonstrates that the results of the algorithm that uses the sample word-emotion lexicon exceed the results of the trained SMO classifier.

Table 9. The weighted average classification results of SMO vs. SF algorithm.

Emotion	Algorithm	P	R	F
Anger	SMO	0.407	0.641	0.498
	SF	0.532	0.885	0.665
Disgust	SMO	0.333	0.253	0.287
	SF	1	0.462	0.632
Fear	SMO	0.648	0.431	0.504
	SF	0.838	0.65	0.732
Happiness	SMO	0.422	0.713	0.53
	SF	0.542	0.888	0.673
Sadness	SMO	0.431	0.272	0.334
	SF	0.731	0.533	0.616
Surprise	SMO	0.48	0.333	0.393
	SF	0.816	0.431	0.564

After that, we checked our assumption that there exists a correlation between the emotion expressed in the sentence and the communication style used by the writer.

4.3.1 Communication style in emotion detection

According to the communication style analysis of the total 1605 tweets shown in table **10**, we can use the exclusion of some emotions upon the presence of a specific communication style.

Table 10. Communication style in 1605 tweets.

	Aggressive	Assertive	Passive
Anger	76.53%	16.91%	6.85%
Disgust	83.33%	9.31%	7.35%
Fear	3.4%	38.11%	58.39%
Happiness	3.83%	86.73%	9.44%
Sadness	32.18%	40.1%	27.72%
Surprise	37.95%	36.75%	25.3%

Looking at the minorities, anger, disgust, and happiness are rarely expressed using the passive communication style. Moreover, few fear and happiness tweets are expressed in the aggressive communication style. In addition, only 9.31% tweets of the disgust emotion category are expressed using the assertive communication style.

The results in table **11** show the issue from communication style point of view.

Table 11. Communication style percentage distribution excluding surprise and sadness.

	Anger	Disg-ust	Fear	Happin-ess
Aggressi-ve	61.86%	33.6%	2.0%	2.5%
Assertive	13.88%	3.88%	26.65%	60%
Passive	11.57%	6.2%	69.0%	13.22%

If we adopt [30] emotional model that suggest that disgust is a secondary emotion of anger, then 95.46% of the aggressive communication style tweets will be expressing anger. Although the assertive communication style is used in the expression of several emotions, it is more associated with the happiness emotion category.

5 Discussion

Our created corpus size is comparable to the OCA corpus size in [23] which consists of 500 reviews, divided into positive and negative ones, in addition to the 1143 corpus in [24] and the 1000 tweets datasets in [18]. Classification algorithms like SMO and NaïveBayes that proved good performance for English text categorization are applicable for Arabic text as well. Similarly to the English language case our experiment demonstrates that SMO classifier outperforms the NaïveBayes classifier for Arabic text categorization by 5.4%. These test results support the claim in the text polarity classification in [23] and [22]. Moreover, the same finding was valid for English language polarity test in [14] that used a linear classifier and in [31]. In the text categorization, the support vectors idea is more effective as it separates the classes with the largest margin and does not depend on the individual words probabilities like the NaïveBayes classifier.

We compared the classification results of our set of tweets using the modified Khoja stemmer and the SMO classifier with the classification results of the 1000 English news headlines in [19]. The results of the comparison concerning the precision and recall results of the different emotions are shown in table **12**. In this table, we refer to the study [19] by the number 1 and to our study with the number 2.

Table 12. P. & R. comparison between our study and [19].

	Study	P	R
Anger	1	0.493	0.265
	2	0.474	0.597
Disgust	1	0.421	0.186
	2	0.488	0.389
Fear	1	0.635	0.437
	2	0.582	0.533
Happiness	1	0.54	0.367
	2	0.623	0.677
Sadness	1	0.525	0.367
	2	0.479	0.345
Surprise	1	0.443	0.292
	2	0.545	0.506

Although the test sets are not similar, and the language used is different, the previous comparison gives a rough estimate of how good the classification performance is.

Moreover, we have added some code that adds our additional features to Khoja stemmer in order to support the Arabic Egyptian dialect and enhance the classification results. It has been shown that our modifications improved the overall classification results by almost 4.4% compared to the original khoja stemmer. Our modified Khoja stemmer solved some of the problems mentioned by Al-Khalifa in [32] such as the negation forms, the extra white spaces, and the spelling mistakes found in the tweets. The preprocessing steps help the classifier to narrow down its feature scope and avoid noise features.

The experiment demonstrates that the preprocessing was more effective in case of limited features where the likelihood of word repetition is higher. Therefore, the classifier can associate words with emotion categories in the training phase and find them frequently in the test phase. Thus, if the training set increases to include tweets that enclose all the emotion related words of a certain language, the preprocessing is supposed to be more effective.

It has been shown that the sample word-emotion lexicon created based on the significant attributes in the tweets enhanced the overall performance of the classification by 22.27% compared to the SMO classification using the train/test option. Comparing this result with [19] and [17], we can support their claim that a word-emotion lexicon would easy the emotion detection task. In [17] the usage of emotion words has positively affected the overall classification performance by 4.35% compared to the blogs training classification performance. Moreover, the emotion related words extracted from microblogs in [19] enhanced the classification results by 19.05 % compared to the original classification.

Yet, the experiment demonstrates that the usage of the SMO classifier and the word-emotion lexicon is a tradeoff. In case of the availability of a dataset with a variety of words, the sample word-emotion lexicon will not be as effective. That is due to the fact that the SMO algorithm uses almost all contained words as features, and it creates the classification model accordingly. This big number of words enriches the model and gives it the ability to deal with random tweets. On the other hand, it has been shown that the usage of a sample word-emotion lexicon has a great benefit if the test data contains these words. However, if such this sample word-emotion lexicon is used to classify a random test set, it will not perform as good as the SMO classifier. Therefore, in order to have the benefit of the lexicon in all cases, a full word-emotion lexicon should be developed and used.

The experiment demonstrates that a reciprocal relationship exists between the emotion and the communication style. It has been shown in section 4.3.1 that a correlation exists between the emotion expressed and the communication style. Based on the analysis of the data, we can exclude the fear and happiness emotions if the communication style is aggressive. Further, we can exclude the disgust emotion if the communication style is assertive or passive. More likelihood should be given to the fear emotion in case of passive communication style. More likelihood should be given to the happiness emotion in case of assertive communication style.

6 Limitations

While analyzing the results of this study, certain limitations need to be taken into account. The dataset that this study focused on is the Egyptian revolution in 2011 topic. Moreover, the size of the dataset (1605 tweets) was not big. However, the annotation of each tweet is very accurate due to two factors. The first one is the doubled number of annotators compared to other studies like [14] and [17]. The second factor is that each tweet underwent more than one annotation run. Moreover, the vocabulary used to express emotions may vary from topic to another according to [19]. Thus, the performance of classification models represented in this study for other topics is not guaranteed. Finally, the study focused on the Arabic language tweets. The preprocessing techniques are customized for the Arabic Language words (standard and slang Egyptian dialect).

7 Conclusion

This study tackled the automatic detection of emotions in textual parts of Twitter the social media website. This process has applications for business development, user interface design, content creation, emergency response, among others. The study was concerned with the capability of classifying Arabic text considering the standard and the slang Egyptian dialect. The dataset under consideration is composed of Arabic text messages collected from Twitter during January/February 2011. The focus topic is labeled with #Jan25 referring to the Egyptian revolution in 2011. A corpus of emotion annotated tweets was developed as well as a sample word-emotion lexicon. Five different preprocessing techniques have been compared. New features were added to Khoja stemmer in order to support some of the slang Egyptian dialect. It has been shown that the use of this modified Khoja stemmer has been associated to the best classification performance. Moreover, the experiment demonstrates that our simple search and frequency algorithm performed better than the SMO classifier in the limited features environment.

We are contributing to the research area of the emotion detection of Arabic content by showing that emotions can be automatically detected from tweets after performing Arabic language related language preprocessing steps. The experiment demonstrates that the preprocessing steps added to Khoja stemmer improved the classification results by 4.4% compared to the original Khoja stemmer performance. In addition, it has been shown that our sample word-emotion lexicon enhances the emotion detection results by 22.27% compared to the SMO classification using the train/test option. Finally, it has been shown that the communication style is closely related to the emotion expressed in case of anger, disgust, fear, and happiness categories. The relationship can be thought of as a reciprocal one.

8 Future Work

The results of this study point to different interesting directions for future work. First, a complete Arabic word-emotion lexicon can be developed. That could be done by considering tweets from different topics and annotating them by emotion. Or, it could be done by going through all the words in the Arabic dictionary and annotating those that have emotion significance.

The WordNet affect, the NRC, and the lexicon extracted from the TEC can be taken as a reference in order to check the inclusion of all corresponding Arabic emotion related words. Moreover, the development of a stemmer that supports the different Arabic slang dialects would be of a great effect. In addition, the expansion of the Arabic tweets corpus with more instances labeled for emotions should be considered. This step would help the training task and make it better for the SMO classifier. Finally, the development of an automatic system that detects the communication style from Arabic text based on the structure of the sentence. This system could facilitate the emotion detection task.

7 REFERENCES

[1] M. S. Thambirajah. 2005. "Psychological Basis of Psychiatry." Churchill Livingstone.

[2] P. Ekman. 1992. "Are there basic emotions. Psychological Review." Psychological Review, volume 99, pages 550–553.

[3] J. A. Russel. 1994. "Is there universal recognition of emotion from facial expression? a review of the cross-cultural studies." Psychological Bulletin, volume 115, pages 102–141.

[4] P. Ekman. 2000. "Basic emotions." In T. Dalgleish and T. Power (Eds.) The Handbook of Cognition and Emotion, pages 45–60. Sussex, U.K.: John Wiley and Sons, Ltd.

[5] Arabic Web Days. 2012. http://www.arabicwebdays.com/. Online; accessed May-2013.

[6] P. Ekman. 1971. "Universals and cultural differences in facial expression of emotions." Lincoln: University of Nebraska Press, volume 19, pages 207–282.

[7] B. Pang and L. Lee. 2008. “Opinion mining and sentiment analysis.” Found. Trends Inf. Retr., volume 2, pages 1–135.

[8] V. Hatzivassiloglou and K. McKeown. 1997. “Predicting the semantic orientation of adjectives.” In Proceedings of the Joint ACL/EACL Conference, pages 174–181.

[9] J. Wiebe. 2000. “Learning subjective adjectives from corpora.” In Proceedings of the Seventeenth National Conference on Innovative Applications of Artificial Intelligence, pages 735-740. AAAI Press.

[10] A. Valitutti, C. Strapparava, and O. Stock. 2004. “Developing affective lexical resources.” PsychNology, volume 2, pages 61-83.

[11] N. Godbole, M. Srinivasaiah, and S. Skiena. 2007. “Large-scale sentiment analysis for news and blogs.” In Proceedings of the International Conference on Weblogs and Social Media (ICWSM).

[12] B. Pang, L. Lee, and S. Vaithyanathan. 2002. “Thumbs up?: Sentiment classification using machine learning techniques.” In Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing (EMNLP ’02), volume 10, pages 79–86, Stroudsburg, PA, USA. Association for Computational Linguistics.

[13] T. Nasukawa and J. Yi. 2003. “Sentiment analysis: capturing favorability using natural language processing.” In Proceedings of the 2nd international conference on Knowledge capture (K-CAP ’03), pages 70–77, New York, NY, USA. ACM.

[14] C. O. Alm. 2005. “Emotions from text: Machine learning for text-based emotion prediction.” In Proceedings of HLT/EMNLP, pages 347–354.

[15] M. Guerini, L. Gatti, and M. Turchi. 2013. “Sentiment analysis: How to derive prior polarities from sentiwordnet.” In EMNLP, pages 1259–1269. ACL.

[16] R. Socher, A. Perelygin, J. Wu, J. Chuang, C. D. Manning, A. Ng, and C. Potts. 2013. “Recursive deep models for semantic compositionality over a sentiment treebank.” In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, pages 1631-1642, Seattle, Washington, USA. Association for Computational Linguistics.

[17] C. Strapparava and R. Mihalcea. 2008. “Learning to identify emotions in text.” In Proceedings of the 2008 ACM Symposium on Applied Computing (SAC’08), pages, 1556-1560, New York, NY, USA. ACM.

[18] H. Saif, Y. He, and H. Alani. 2012. “Semantic sentiment analysis of twitter.” In International Semantic Web Conference (1), pages 508–524.

[19] S. Mohammad. 2012. “Emotional tweets.” 2012. The First Joint Conference on Lexical and Computational Semantics, Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval 2012), pages 246–255. Association for Computational Linguistics.

[20] D. F. Gurini, F. Gasparetti, A. Micarelli, and G. Sansonetti. 2013. “A sentiment-based approach to twitter user recommendation.” In RSWeb@RecSys.

[21] M. Al-diabat. 2012. “Arabic text categorization using classification rule mining.” In Applied Mathematical Sciences, volume 6, pages 4033-4046.

[22] B. Al-Shargabi,W. Al-Romimah, and F. Olayah. 2011. “A comparative study for arabic text classification algorithms based on stop words elimination.” In Proceedings of the 2011 International Conference on Intelligent Semantic Web Services and Applications (ISWSA ’11), pages 1–5, New York, NY, USA. ACM.

[23] M. Rushdi-Saleh, M. T. Martn-Valdivia, L. A. Lpez, and J. M. Ortega. 2011. “Bilingual experiments with an arabic-english corpus for opinion mining.” In Galia Angelova, Kalina Bontcheva, Ruslan Mitkov, and Nicolas Nicolov, editors, RANLP, pages 740–745. RANLP 2011 Organising Committee.

[24] A. El-Halees. 2011. “Arabic opinion mining using combined classification approach.” International Arab Conference on Information Technology (ACIT’2011), volume26, Riyadh, Saudi Arabia.

[25] T. H. Soliman, A. Hedar, M. Ali, and M. M. Doss. 2013. “Mining social networks’ arabic slang comments.” In Proceedings of IADIS European Conference on Data Mining 2013 (ECDM’13).

[26] A. F. El-Gohary, T. I. Sultan, M. A. Hana, M. M. El-Dosoky. 2013. “A Computational Approach for Analyzing and Detecting Emotions in Arabic Text.” In International Journal of Engineering Research and Applications (IJERA), volume 3, pages 100-107.

[27] Apache Software Foundation. 2010. http://lucene.apache.org/core/old_versioned_docs/versions/3_0_3/api/contrib.-analyzers/org/apache/lucene/analysis/ar/ArabicAnalyzer.html. Online; accessed June-2013.

[28] S. Khoja. 2001. http://zeus.cs.pacificu.edu/shereen/research.htm. Online; accessed June-2013.

[29] Arab geeks. 2009. Arabic stop words. http://arabicstopwords.sourceforge.net/. Online; accessed April-2013.

[30] W. G. Parrott. 2001. “Emotions in Social Psychology: Essential Readings.” Key readings in social psychology. Psychology Press.

[31] T. Joachims. 1998. “Text categorization with support vector machines: Learning with many relevant features.” In European Conference on Machine Learning (ECML), pages 137-142, Berlin. Springer.

[32] L. Albraheem and H. S. Al-Khalifa. 2012. “Exploring the problems of sentiment analysis in informal arabic.” In Proceedings of the 14th International Conference on Information Integration and Web-based Applications (IIWAS’12), pages 415–418, New York, NY, USA. ACM.