International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. #, No.# (Mon-20..)

Sentiment Analysis of Palestinian Arabic Dialect Using Lexicon-Based Approach

Mohammed KH. Zoroub[1], Ashraf Y. Maghari[2] and Abdelkareem M. Alashqar[3]

[1,2,3]Faculty of Information Technology, Islamic University of Gaza, P.O. Box 108, Gaza, Palestine

E-mail address: [1]appeappe8@gmail.com, [2]amaghari@iugaza.edu.ps, [3]aashgar@yahoo.com

Received ## Mon. 20##, Revised ## Mon. 20##, Accepted ## Mon. 20##, Published ## Mon. 20##

Abstract: Most Arab users express their opinions on social media platforms using colloquial dialect which contains a huge quantity of unstructured and ambiguities data. These data require treatment by utilizing sentiment analysis (SA) techniques to discover opinions polarity that deemed useful for stakeholders. In Palestinian dialect there are many problems related to their nature, such as abbreviations and lack of standardization rules for grammar and spelling. These issues made the extract opinions process very challenging task in the SA area. In this paper, a rule-based sentiment lexicon for Palestinian dialect is proposed with novel rules and concepts to classify users' comments into positive, negative, and neutral. Also, a grouping-terms technique is proposed in the preprocessing step to outdo the issues related to texts in Palestinian dialect such as writing different spellings (shapes) that have one meaning for the same word. Additionally, a polarity lexicon for Palestinian dialect has been created during the grouping-terms process. The proposed lexical-classifier achieved better results when using the grouping-terms technique instead of stemming. It achieved an accuracy of 85% when using two classes, and an accuracy of 80.5% when using three classes, which is considered a very good performance in SA approaches. Our results showed that the development of rules and polarity lexicon for Palestinian dialect terms can be considered as a good implication for further related studies.

Keywords: Palestinian dialect, Sentiment Analysis, Lexical Approach, Grouping-Terms, ISRI stemmer

1. Introduction

Currently, the proliferation of social media platforms with a mass of users across the Arab world led to the presence and abundance of textual data. These data consist of useful information such as user's opinions towards certain topics. In order to extract those opinions, sentiment analysis (SA) methods should be used. The SA has become a popular task for the natural language processing (NLP) [1], its target is to determine whether the given textual existence is subjective or objective, and to recognize the polarity of subjective texts. In State of Palestine, Facebook is considered as the most popular social media platforms as shown in Figure 1, where the number of Facebook users in Palestine reached 2,844,000 users in May 2020, which accounted for 52.8% of its entire population [2]. Due to our approach, we opted Facebook as the main source of textual data.

Evidently, the Arab's social media users tend to write their comments by colloquial dialect, which always varies from country to country and from culture to another. The differentiation of dialects can also exist in the same country. In addition, Arabic dialects are grouped into six main categories based on Arabic regions which are: Maghreb, Egyptian, Gulf, Iraqi, Levantine, and Yemeni [3]. Levantine also called SHAMI which is the dialect spoken in Palestine, Syria, Jordan and Lebanon countries. In our work, we will focus on Palestinian dialect, which is considered as part of the South Levantine Arabic dialect sub-group [4].

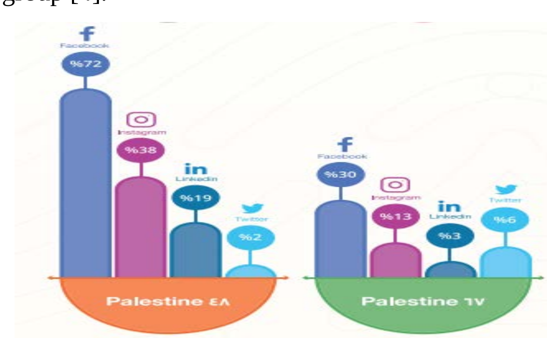

Figure 1. Percentage of Facebook users in the State of Palestine

E-mail: aashgar@yahoo.com

Undoubtedly, Arabic dialects deemed more complex than Modern Standard Arabic (MSA) in terms of morphology and orthography. The challenges associated with the Arabic dialects contents in social media can be summarized as follows [3][4][5][6][7][8]:

- Written as unstructured language.
- Include spelling mistakes.
- Use slangs and abbreviations.
- Use vernacular expressions and proverbs.
- Use emoticons symbols in writing.
- Include repeated letters.

In the same context, textual data written by Palestinian dialect suffer from several problems such as:

- A term can be written in different shapes that have the same meaning, for example: "sweety" "حلو" potentially written as: " حلو, الحلو, حلوين, ما احلاه, شو هالحلي ".
- There are no standardization rules as use negation to confirm positive meaning like: " ما اروعك, ما احسنك, ما احلاك", whereas, in other situation, the same negation is used to confirm negative meaning like: "ما حبيت, ما عجبني, مافهمت".
- Use different pattern of the negation, for example, append "chen" "ش" character at the end of the word, for example: "بحبش, بفهمش, بعرفش".

Two main approaches can be utilized in SA to mine user's opinions which are machine learning (ML) approach and lexicon-based approach where both of them can be combined to produce a hybrid one. ML approach depends on constructing classification model with training labeled data whereas lexicon-based approach depends on creating lexicon for terms of language with polarity scores [9]. In ML approach the training data is essential for the classifier to predict the sentiment of text, while the training process is not needed for lexicon-based approach because it depends on the prepared translated lexicon in determining the sentiment that embedded in the text.

Additionally, the ML approach can be widely used in SA for Modern Standard Arabic (MSA) [3] because its structure and grammar are almost stable over time. And the labelling, annotation and patterns recognition used in SA for MSA can be simplified and automated through the iterative learning process of ML. Whereas, the Arabic dialect has high diversity of written forms and considerable number of different negation shapes among Arab nations and regions, which makes it hard to achieve the automation and learning processes of ML. So, the lexicon-based approach still provides powerful techniques for SA in Arabic dialect such as handling negation, polarity scoring, feature extraction and meaning of text.

Although there are existing research works in SA for Arabic dialect using ML and lexicon-based approaches, these efforts suffer from several deficiencies such as lack of resources, small lexicon terms and less accuracy [10]. Furthermore, there is a little attention on providing lexicon for Levantine dialect especially for Palestinian dialect [11].

Unlike the existing approaches, our work focuses on improving the performance of the lexicon-based approach by proposing novel rules, which have the ability to handle several linguistic features in the Palestinian dialect as negation, boost-increased sentiment, boost-decreased sentiment, and modified sentiment. Also, we propose a new technique called "grouping-terms" in preprocessing step which is applied instead of the stemming. The main concept behind the proposed technique is to reduce the number of word spellings which have polarities into one possible form. For this purpose, a new lexicon containing these word forms is created. Our main contribution in this work are concluded as follows:

- Create a sentiment lexicon for texts in Palestinian dialect.
- Propose novel rules for handling some linguistic features in Palestinian dialect.
- Propose a novel technique called "grouping-terms" to overcome the problems related to text written in Palestinian dialect.

This paper is organized as follows: Section 2 provides literature review for SA approaches that conducted on Arabic dialects. Our proposed methodology for handling Palestinian dialect is presented in Section 3. Section 4 presents the experiments and discusses the results. Finally, Section 5 concludes the paper and suggests future works.

2. Related Works

Sentiment analysis (SA) for Arabic dialect is considered as main interest of many researchers. Some researchers followed the ML approach to achieve SA for Arabic dialect. For instance, the studies in [12][13] performed SA for general Arabic dialect while the studies in [6][14] provided SA for Saudi dialect, the studies in [15][16] achieved SA for Jordanian dialect, the work in [17][18] handled the SA for Sudanese dialect, and recently the authors in [19] provided SA for Algerian dialect with transformers.

The authors in [20] introduced a literature review for 60 selected studies that handled SA for Arabic dialect focusing on using various techniques of ML. A recent study [3] provided a more comprehensive systematic literature review for 72 selected research articles for using SA in Arabic dialect where the authors handled both ML and lexicon-based approaches in their review. They argued that the lexicon-based approach still provides considerable capabilities in SA for Arabic dialects.

Focusing on the research articles that adopted the approach of lexicon-based for Arabic dialect, the authors in [21] proposed a new sentiment lexicon for Saudi dialect called SauDiSenti that included 4431 words and phrases from modern standard Arabic (MSA) and Saudi dialect, which have been extracted manually from Saudi dialect twitters. Their lexicon contains 24% and 76% of negative and positive entries respectively. In the experimental side, their lexicon showed promising results in terms of accuracy in comparison to AraSenTi (a larger Arabic sentiment lexicon) [22] when they used positive, negative, and neutral classes, whereas the AraSenTi has outperformed SauDiSenti in terms of accuracy when using only two classes; positive and negative.

The researchers in [6] built a large-lexicon for the Saudi dialect and developed a weighted lexicon-based algorithm to find relations between polarity and non-polarity words, then weighted these words based on their associations. Further, the authors proposed novel rules for treating some linguistic features such as negation and supplication. According to their experiments, their algorithm achieved the best results on Saudi twitter dataset [23] and Egyptian dataset [5]. Also, the authors in [24] build lexicon for achieving sentiment analysis for Tunisian dialect by fetching the lexicon about the features needed for trained supervised classifier. The showed that this approach is effective and reached best performance results.

On the other hand, some research studies have been achieved on Egyptian dialect. For instance, the author in [25] proposed an improved version of the sentiment lexicon called NileULex that pre-created by [5] and [26]. The authors developed the lexicon through manual additions where the lexicon size was close to 6000 terms and compound phrases. Their results showed that the improved lexicon outperformed the system that used uni-gram, bi-gram, and TF-IDF weights in terms of accuracy. It also outperformed the lexicon system that used translation approach such as EmoLex [27]. In the same context, the authors in [28] continued their efforts and presented a new approach for extending an Arabic sentiment lexicon automatically. To avoid constructing a lexicon from scratch, they have been used NileULex [25] as an initial terms-lexicon. Their experiments showed that the SA using the lexicon entries are more accurate than sentiment lexicon entries obtained using ML or distant supervision methods. It was noted that these authors have used the stemming technique for a minimized number of polarity features in their lexicons. Unfortunately, according to [6][29], this technique showed the weakness in handling with Arabic dialects.

However, our work focuses on Palestinian dialect, which is considered as part of the South Levantine Arabic dialect sub-group [4]. Our proposed approach applied a new technique in the preprocessing step called “grouping-terms” in order to overcome some issues related to the Palestinian dialect. Also, we proposed novel rules for treating some of the linguistic features in the dataset such as negations, increased sentiment, decreased sentiment, and modified sentiment.

Some researchers presented a polarity lexicon for the Palestinian dialect [8], which included 2027 positive tweets, 3289 negative tweets and 777 neutral tweets. They used these tweets as a dataset for applying two types of supervised learning models for text in Palestinian dialect. The first type of learning models is for basic binary text classification models. These basic models were generated using two popular binary classifiers: Support Vector Machine (SVM) and Naïve Bayes (NB). The second learning model aims to improve the accuracy of the basic learning models by extending the feature space with features extracted from the constructed Polarity Lexicon. Their results demonstrate that the SVM algorithm achieved the best accuracy, which is 91.78%. The authors in [4] presented the first annotated corpus for the Palestinian Arabic dialect called "Curras" with over 55,960 tokens annotated with rich morphological and semantic information. Recently, the authors in [30] developed the ArSenTD-Lev which is a corpus for SA in Arabic Levantine tweets. They annotated each tweet based on manual analysis, their annotations rely on the objective to which the sentiment was expressed, how the sentiment was expressed, and the topic of the tweet. Their corpus consists of 4,000 tweets collected from Levantine countries.

The approaches in [8] and [30] depend on the corpus-based method, which requires to annotate each given sentence manually. In contrast, "Curras" [4] focused on annotating every word with a set of metadata attributes and developed a set of spelling rules for the Palestinian dialect. However, part of our approach aimed to build a rule-based sentiment lexicon for text in Palestinian dialect, which be able to classify the texts into positive, negative, and neutral. In addition, due to the lack of available resources related to the Palestinian dialect, we intended to create a polarity lexicon for Palestinian dialect from scratch.

3. Scope and Methodology

In this section our proposed methodology is provided in order to achieve the objective of this study. This methodology consists of five main steps: dataset collection and preprocessing, process the dataset using the “Grouping-Terms” technique, create polarity lexicon, create sentiment lexicon and evaluation. Figure 2 demonstrate the main steps of our proposed methodology.

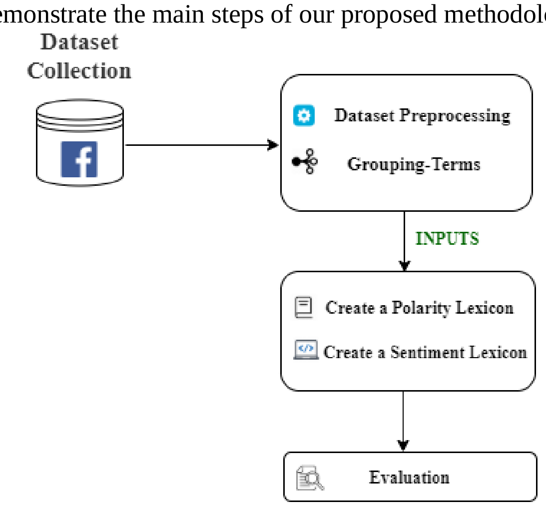

Figure 2. The main steps of the proposed methodology

A. Data Collection and Preprocessing

Initially, we collected our datasets from Facebook using NETVIZZ Facebook API to fitch the content of comments included in Facebook pages that located only in Palestinian regions where the comment incudes up to 150 characters. Owing to our work interested only in a sentence level, the total number of collected dataset reached 3870 comments. On the other hand, several techniques of NLP have been applied in preprocessing data which includes tokenization, removing repeated letters, normalization and removing stop words. Then we proposed the grouping-terms technique and applied it on the dataset.

There are several algorithms of stemming were developed for Arabic language [7], such as Khoja-stemmer, light-stemmer and ISRI Arabic stemmer. The former and the latter algorithms are considered as intensive stemmers which rely on removing affixes, and altered some letters in the word to extract the root of it. Whereas the second stemmer is deemed more simply, since it does not need any grammatical analysis to get the root [29]. Recently, many efforts [6][21][28][29] in Arabic sentiment analysis have been carried out by utilizing one or more of these stemmers as a step of preprocessing text data. Some of these efforts [28][29] have concluded that, the root-based stemmer often gives us incorrect root, which is convenient to our hypothesis that the stemmer is compatible less with textual data written by Arabic dialects. On the other hand, we used the online Arabic stemmer [31][32] in order to test our hypothesis. Figure 3 and Figure 4 show the results of our testing, which apparently show the weaknesses of the Arabic stemmers in handling the text of Palestinian dialect. Neither ISRI-stemmer nor light-stemmer are able to return the correct stem or to minimize the number of word forms, which have polarities, into one possible form.

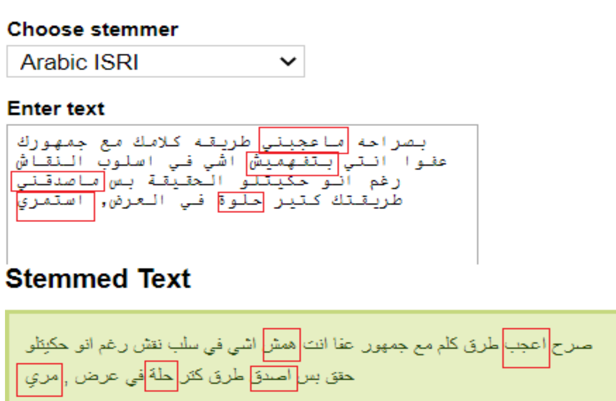

Figure 3. ISRI Stemmer

Figure 4. Arabic Light Stemmer

B. The Proposed "Grouping-Terms" Technique

The Palestinian users of social media unconcerned to discipline with a linguistic spelling approach for tuning their textual comments. They write many spellings for the same word and they use the same negation tool either to boost positive meaning or to boost negative meaning. In order to outpace these issues, we proposed the "grouping-terms" technique which configured with Palestinian dialect, to reduce and uniform the polarity features in our dataset. Our proposed technique aims to enhance the accuracy of the opinion mining process. The implementation of this technique concludes in several steps:

- Firstly, we created a dictionary that has a list of polarity terms, Table 1 shows a sample of the polarity terms dictionary. This dictionary includes 930 polarity terms which made up into 500 positive terms and 430 negative terms.

- Secondly, we search for all different word shapes written for each term in our dataset. Sequentially, the results are assigned to the target word, this step is done for each term in the list of our dictionary.

- Finally, we replace each polarity term in our dataset preprocessed by their equivalents in our dictionary. Table 2 shows the results of applying grouping-terms. It is noteworthy that all aforementioned steps have been done programmatically by executing a piece of code in Python programming language.

Table 1. Sample terms of our polarity terms dictionary

Terms	Different Shapes of Each Term
'حلو'	[‘ما احلاه ',' احلا ',' احلي ',' احلاه ',' محلاكم ',' حلوين ', ' مااحلاهم ',' احلاهم ',' حلوه ' , ' احلاكي ' , ' حلوات' ..الخ]
'رائع'	[‘روعه ',' ماروعك ',' اروعك ',' رائعين ', ' شورائع ',' شوروعه ',' اروعكم ' , ' مااروعكم' .. الخ]
'حلل'	['الاحتلال ', ' لاحتلال ', ' احتلال ', ' احتلينا ', ' احتلالنا ', ' لاحتلالنا ',' احتلتنا ' , ' محتله ', ' المحتل ', ' المحتله ' , ' المحتلين']
'بحبش'	['محبيتش ', ' ماحبيتها ', ' بحبهوش ' , ' ماحبيت ', ' ماحبيتو ' , ' مبحب ' , ' منحبكاش ', ' بنحبش ',' وبنحبش']

Table 2. Results of applying the grouping-terms technique

Comments	Apply the Grouping-Terms
علي فكرة استاذ احمد كتير حبيتها للروايه	علي فكرة استاذ احمد كتير حبب للروايه
شو هالروعه شو ابدعتوا ربنا معاكم ويوفقكم	شو رائع شو بدع ربنا معاكم وفق
اختياراتك كل مالها بتصير من السيئ الي الاسوا	اختياراتك كل مالها بتصير من سيئ الي سيئ
انت بتتهزا علي الصيام الدين مش مسخرة	انت هزأ علي الصيام الدين مش سخر

C. Create a Polarity Lexicon

The polarity lexicon for Palestinian dialect has been created during the grouping-terms process, which means that we passed each polarity term (word) that exist in a dictionary (which presented in subsection B) into a polarity lexicon. Also, we scored each term manually from 1 to 4 for positive terms and from -4 to -1 for negative terms. The number 1 indicate positive term and 4 indicate extremely positive term, whereas -1 indicate negative term and -4 indicate extremely negative. Table 3 shows the examples of our polarity lexicon.

Table 3. Examples of positive and negative terms of Polarity Lexicon

Polarity Terms	زنخ	سخيف	زعج	فخر	حبب	جدع
Score	-4	-2	-1	2	1	1

D. Create the Rule-Based Sentiment Lexicon

In this section, we explain in detail the rules and conditions that were implemented to create a sentiment lexicon for text in Palestinian dialect. The entity behind this implementation is represented in providing the lexical-classifier, which is able to handle a multi-class classification problem, where each comment is labeled as positive, negative or neutral. In the same context, we worked on designing the rule-based sentiment lexicon that contains programmed components to detect and treat some frequented words in the dataset which affect semantic orientation drastically, we categorized them into four types as shown in Table 4.

Table 4. Types of influence words

Types of words	Examples
Negation	لا , مش , مو , ما , غير , بدون, فش
Increasingly sentiment	جدا , كثير , كتير , قمه , بشده
Decreasingly sentiment	اقل , قليل , انعدام , قله
Modifiers	لكن , بس

To demonstrate how lexical-classifier works, we used the flow-chart shown in Figure 5, to illustrate the functions and steps involved in the classifier. Initially, we plugged the polarity lexicon which described in subsection C into our lexical-classifier and converted it into dictionary format. Then, the scored polarity features are used to assist in the sentiment classification task. After that, the programmatic functions were implemented to achieve our sentiment rules, such as negation-check, increase-sentiment-check, decrease-sentiment-check, and modifiers-check, which we explore in later subsections. Thereafter, the dataset is loaded into a lexical-classifier with tokenizing each word belong to every single comment, whereas, each comment will pass to our rules check to identify their polarity. Finally, similar to [29], positivity or negativity words are aggregate for each comment to sum their scores.

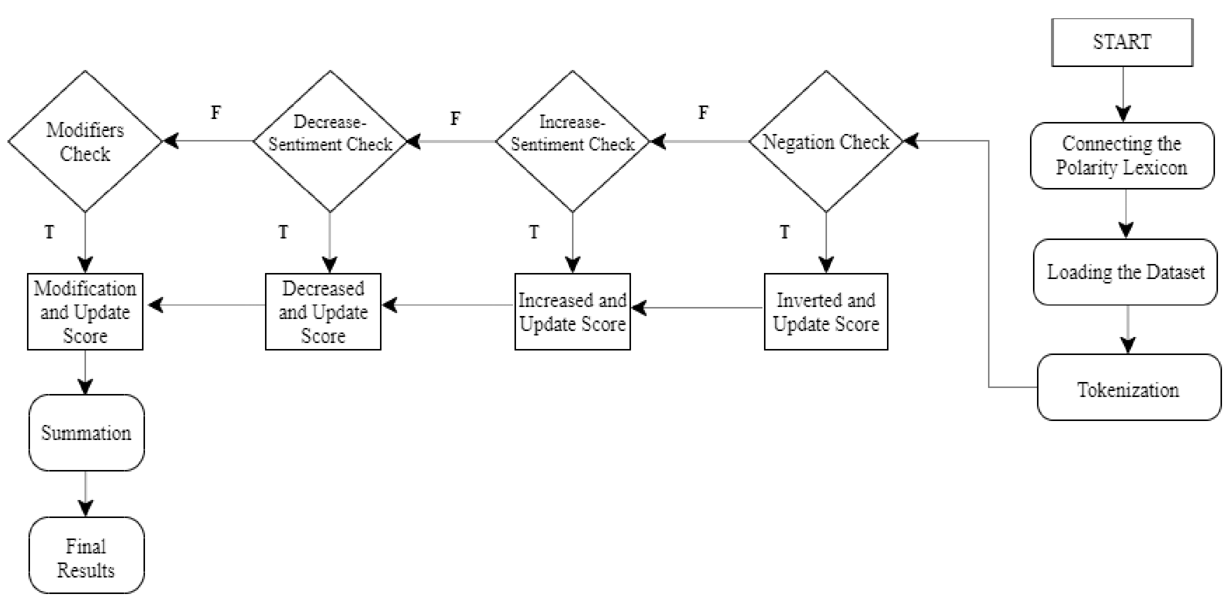

Figure 5. Flowchart of processing rules in the lexical-classifier

International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. #, No.# (Mon-20..)

On the other hand, our sentiment lexicon is considered unsupervised classifier, which consists of the following rules for predictions:

- **Negation-Check:** This rule aims to determine if the input text contained negation words, and check if one or two words are preceding the polarity term, then the polarity score for this term is inverted by multiplying its score by -1. Figure 6 shows an example code of negation rule.
- **Increase-Sentiment-Check:** This rule aims to check if increase- sentiment words preceding or following the polarity term, then increasing the polarity score for this term, by multiplying its score by 1.5.
- **Decrease-Sentiment-Check:** Also, to check if decrease- sentiment words preceding or following the polarity term, then decreasing the polarity score for this term, by multiplying its score by 0.5.
- **Modifiers-Check:** This rule aims to check if modifier words are existing in input text as contrastive conjunction, then the polarity score of the term before the modifier word is multiplied by 0.5 and the polarity score of term that follow the modifier word is multiplied by 1.5. Figure 7 shows an example code of the Modifiers-Check rules.

```
if (i > 0 and words [i - 1] == "مش") \
or (i > 1 and words [i - 2] == "مش") \
or (i > 2 and words[i - 3] == "مش" \
and words [i - 1] in [ "ولا"،"ولا حتي"،"او"، "ولاحتي"] ):
        valence = self.lexicon[item] * -1
```

Figure 6. Code example of Negation-Check

```
if 'بس' in words():
        bi = words.index('بس')
        for sentiment in sentiments:
            si = sentiments.index(sentiment)
            if si < bi:
                sentiments.pop(si)
                sentiments.insert(si, sentiment * 0.5)
            elif si > bi:
                sentiments.pop(si)
                sentiments.insert(si, sentiment * 1.5)
    return sentiments
```

Figure 7. Code example of Modifiers-Check

Table 5 shows 7 examples of applying our grouping-terms technique and rules on the comments. For example, the class of the comment in the first row is determined as follows: Firstly, our sentiment lexicon check if the given comment has a polarity words, and return their original scores, in this case the function-based lexicon found one polarity word (shame=عيب) and return -1. Then, the negation-check rule was activated owing presence the negation word in this comment (not=مش). So, the negation rule inverted the polarity score for this word by multiplying its score by -1, and became +1 as shown in Table 5 (Row 1). The final score is +1, which means that the comment class in the first row is positive. Another example is determining the comment class in row number six as follows: After capturing two polarity words as (problem=مشكله) and (benefit=ربح) in the comment, the decrease-sentiment-check rule was activated owing presence the decrease word (less=اقل). So, the decrease-sentiment rule dropped the polarity score of (benefit=ربح) by multiplying its score by 0.5, and became +0.5 as shown in Table 5 (Row 6). The final score is (-1(problem=مشكله) + 0.5(benefit=ربح)) = -0.5, which means that the comment class in the six row is negative.

E. Evaluation

The performance of our sentiment lexicon is evaluated in term of accuracy. We first evaluate our sentiment lexicon by using the dataset, on which the proposed grouping-terms has been applied. Second, the sentiment lexicon is evaluated by using the dataset on which the Arabic ISRI stemmer has been applied. Noteworthy, the different number of classification classes (two and three) have been used in each phase. The lexicon is valuated in term of accuracy measurement. In our case, accuracy of a sentiment lexicon refers to the result of the overall measured values that are correctly classified by the sentiment lexicon. The main equation of the accuracy is as following:

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN)$$

Where TP (True Positive) and TN (True Negative) mean that the classifier correctly predicts the positive class and negative class respectively. Whereas FP (False Positive) and FN (False Negative) mean that the classifier incorrectly predicts the positive class and negative class respectively.

E-mail: aashgar@yahoo.com

International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. #, No.# (Mon-20..)

Table 5. Examples of applying the grouping-terms technique and rules

<table>
<tr><th>#</th><th>Comments</th><th colspan="2">Original Score</th><th colspan="2">Apply the Lexical-Classifier</th><th>Final Score</th><th>Class</th></tr>
<tr><td rowspan="2">1</td><td rowspan="2">مش عيب انك تشتغل وانت بمرحلة الدراسة</td><td colspan="2">عيب</td><td colspan="2">مش عيب</td><td rowspan="2">1</td><td rowspan="2">Positive</td></tr>
<tr><td colspan="2">-1</td><td colspan="2">1</td></tr>
<tr><td rowspan="2">2</td><td rowspan="2">حضرتك مش فاهم ولا حتي عارف</td><td>فهم</td><td>عارف</td><td>مش فهم</td><td>ولا حتى عارف</td><td rowspan="2">-2</td><td rowspan="2">Negative</td></tr>
<tr><td>1</td><td>1</td><td>-1</td><td>-1</td></tr>
<tr><td rowspan="2">3</td><td rowspan="2">حلقة اليوم سيئة جدا</td><td colspan="2">سيئ</td><td colspan="2">سيئ (جدا)</td><td rowspan="2">-3</td><td rowspan="2">Negative</td></tr>
<tr><td colspan="2">-2</td><td colspan="2">-3</td></tr>
<tr><td rowspan="2">4</td><td rowspan="2">انتو ناس رائعين جدا</td><td colspan="2">رائع</td><td colspan="2">رائع (جدا)</td><td rowspan="2">3</td><td rowspan="2">Positive</td></tr>
<tr><td colspan="2">2</td><td colspan="2">3</td></tr>
<tr><td rowspan="2">5</td><td rowspan="2">انا مع الهدنه طويلة الامد لانها اقل خسارة</td><td>مع</td><td>خسر</td><td>مع</td><td>خسر</td><td rowspan="2">0.5</td><td rowspan="2">Positive</td></tr>
<tr><td>1</td><td>-1</td><td>1</td><td>-0.5</td></tr>
<tr><td rowspan="2">6</td><td rowspan="2">مشكلة الوظيفه الحكوميه بدها شغل ووقت و ربح اقل</td><td>مشكلة</td><td>ربح</td><td>مشكلة</td><td>ربح</td><td rowspan="2">-0.5</td><td rowspan="2">Negative</td></tr>
<tr><td>-1</td><td>1</td><td>-1</td><td>0.5</td></tr>
<tr><td rowspan="2">7</td><td rowspan="2">بشوف طاقات رائعه بالتمثيل بس توظيفها سيئ</td><td>رائع</td><td>سيئ</td><td>رائع</td><td>سيئ</td><td rowspan="2">-0.5</td><td rowspan="2">Negative</td></tr>
<tr><td>2</td><td>-1</td><td>1</td><td>-1.5</td></tr>
</table>

4. EXPERIMENTAL RESULTS AND DISCUSSION

Our experiments aim to compare the performance of a sentiment lexicon in terms of accuracy, one time using a stemming technique (ISRI) and the other time by using our proposed Grouping-Terms technique.

Similar to [33], 900 comments are picked from our dataset (3870 comments), and manually annotated into 350 positive comments, 350 negative comments and 200 neutral comments. We evaluated the two sentiment lexicons (ISRI, Grouping-terms) by comparing it with our handmade annotated comments.

A. Sentiment lexicon using Arabic ISRI Stemmer

The ISRI stemmer technique was applied to the dataset, which is 900 comments. To achieve this experiment, several steps have been conducted which shown in Figure 8.

Firstly, the preprocessing steps were applied to the dataset. Secondly, the polarity terms were extracted manually from the dataset, then a stemming (ISRI stemmer) technique is applied to them. After that, if the word stemming is correct, then the polarity term is added manually to the polarity lexicon, else the term is discarded. Posteriorly, the rules of the sentiment lexicon were applied on the dataset comments. Finally, the lexicon is evaluated using accuracy measurement. Examples of polarity terms added to the polarity lexicon or discarded using ISRI stemmer, are concluded in Table 6.

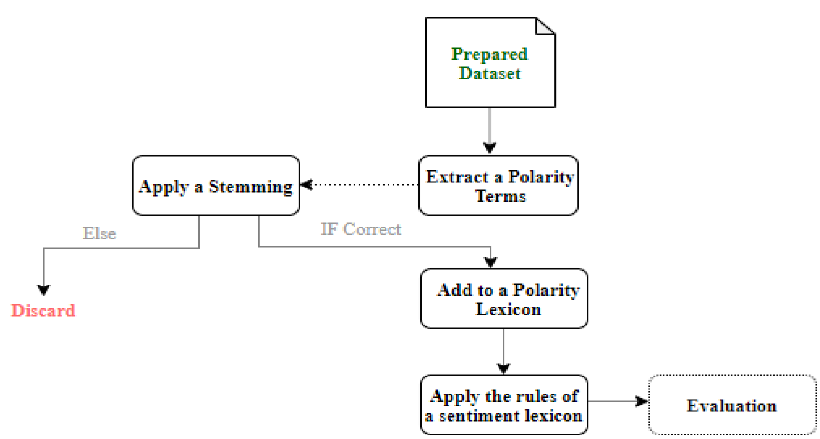

Figure 8. The main steps of experiment of ISRI stemmer

E-mail: aashgar@yahoo.com

Table 6. Examples of terms added to the Polarity lexicon or discarded using ISRI stemmer

Comments	Polarity Terms	Class	Stemming	Add to a Polarity Lexicon
بصراحه ماعجبني طريقه كلامك مع جمهورك	ماعجبني	Negative	اعجب	discarded
عفوا انتي بتفهميش اشي في اسلوب النقاش	بتفهميش	Negative	همش	discarded
رغم انو حكيتلو الحقيقة بس ماصدقني	ماصدقني	Negative	اصدق	discarded
طريقتك كتير حلوة في العرض, استمري	حلوة	Positive	حلو	Added to polarity lexicon
طريقتك كتير حلوة في العرض, استمري	استمري	Positive	مري	discarded

1) Experimental Results Using ISRI Stemmer

In case of using a binary class (negative and positive), we obtained 200 comments that were incorrectly classified by the sentiment lexicon; 80 positive comments, and 120 negative comments. This means that 270 positive comments and 230 negative comments are correctly classified. The classification accuracy can be obtained using the following formula:

Accuracy = (270 + 230) / (270 + 230 + 80 + 120) = 71%

While, in the case using the multi-classification classes (positive, negative, neutral), the results obtained from a comparison process have been filled in a confusion matrix table (Table 7). The correct classifications (true class = predicted class) are located on the shaded diagonal from top left to bottom right, while the other values are the false classifications, as shown in Table 7. So, Accuracy is the sum of the diagonal divided by the total. The classification accuracy is determined by the following formula:

Accuracy = (230 + 270 + 107) / (230 + 270 + 107 + 120 + 35 + 80 + 25 +13 +20) = 67%

Table 7. The Results of sentiment lexicon with ISRI stemmer for 3 classes

	True Negative	True Positive	True Neutral
Predicted. Negative	230	120	35
Predicted. Positive	80	270	25
Predicted. Neutral	13	20	107

B. Sentiment Lexicon Using Grouping-Terms

In these experiments, the preprocessing steps including the proposed Grouping-Terms technique were applied to the same dataset used in subsection A. Then, the rules of the sentiment's lexicon are applied. Finally, the lexicon is evaluated using accuracy measurement.

1) Results of Using the Grouping-Terms

In the case of using a binary class, 100 comments were incorrectly classified by the sentiment lexicon; 40 positive comments and 60 negative comments. This means that 310 positive comments and 290 negative comments are correctly classified.

Accuracy = (310 + 290) / (310 + 290 + 60+ 40) = 85%

While, in the case of using the multi-classification classes, the results obtained were located in Table 8. So, Accuracy is the sum of the diagonal divided by the total. The classification accuracy is determined by the following formula:

Accuracy = (290 + 310 + 125) / (290 + 310 + 125 + 60 + 35 + 40 + 25 + 5+10) = 80.5%

Table 8. The results sentiment lexicon with Grouping-Terms for 3 Classes.

	True Negative	True Positive	True Neutral
Predicted. Negative	290	60	35
Predicted. Positive	40	310	25
Predicted. Neutral	5	10	125

C. Results Discussion

We noticed that the sentiment lexicon achieved the highest performance in terms of accuracy when we applied the proposed Grouping-Terms on the dataset. The highest accuracy achieved is 85% when using a binary classification while coming next on 80.5% accuracy when using the multi-classification classes. These results come down to 71% and 67% respectively, when applied a stemming technique (Arabic ISRI), as shown in Figure 9. The main reason for these results is that the stemming technique is not configured to handle Arabic colloquial dialects like Palestinian dialect, similar to [28][29]. This means that many polarity terms written in Palestinian dialect do not have the correct root. In addition, many polarity terms were not added to a polarity lexicon because they did not give the appropriate meaning when stemming them, as annotated in Table 6.

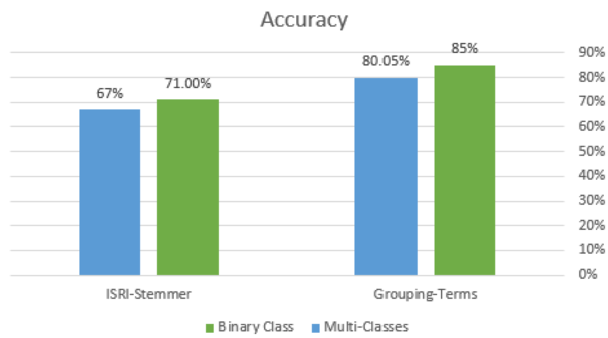

Figure 9. The Results of sentiment lexicon using Grouping-Terms and Arabic ISRI

5. CONCLUSION and Future Work

In this paper we presented a lexicon-based approach in sentiment analysis for the Palestinian dialect. Novel rules based on sentiment lexicon were proposed to classify text in Palestinian dialect into Positive, Negative, and Neutral opinions. Also, we proposed grouping-terms technique which applied to the dataset as an alternative of stemming. The grouping-terms technique overcome the problem of writing different spellings that have one meaning for the same word Palestinian dialect. Our sentiment lexicon has been evaluated in term of accuracy. We obtained an accuracy of 85% and 80.5% for two classes and three classes respectively when applying the grouping-terms technique on the dataset. Whereas these scores came down to 71% and 67% respectively when applying the stemming technique. This means that using the grouping-terms technique made the sentiment analysis more accurate and achieved efficient and satisfied results.

An extension to this work can be achieved by expanding our dictionary of grouping-terms and polarity lexicon in order to improve the accuracy of sentiment lexicon of text in Palestinian dialect. Where publishing of such polarity lexicon for public used is needed. Also adding other sentiment rules that might be discriminative such as emoticons and sarcasm can be performed. A hybrid approach that combines lexicon-based with ML techniques can be utilized.

REFERENCES

[1] M. Al-Ayyoub, A. A. Khamaiseh, Y. Jararweh, and M. N. Al-Kabi, “A comprehensive survey of arabic senti-ment analysis”, Inf. Process. Manag., vol. 56, no. 2, pp. 320–342, 2019.

[2] Socialstudio, “Digital and Social media report in Pales-tine,” 2019.

[3] Y. Matrane, F. Benabbou and N. Sael, “A systematic literature review of Arabic dialect sentiment analysis”, Journal of King Saud University – Computer and Information Sciences 35, 101570, 2023.

[4] M. Jarrar, N. Habash, F. Alrimawi, D. Akra, and N. Zalmout, “Curras: an annotated corpus for the Palestin-ian Arabic dialect,” Lang. Resour. Eval., vol. 51, no. 3, pp. 745–775, 2017.

[5] S. R. El-Beltagy and A. Ali, “Open issues in the senti-ment analysis of Arabic social media: A case study,” in 2013 9th International Conference on Innovations in Information Technology (IIT), pp. 215–220, 2013.

[6] A. Assiri, A. Emam, and H. Al-Dossari, “Towards en-hancement of a lexicon-based approach for Saudi dia-lect sentiment analysis,” J. Inf. Sci., vol. 44, no. 2, pp. 184–202, 2018.

[7] M. G. Syarief, O. T. Kurahman, A. F. Huda, and W. Darmalaksana, “Improving Arabic Stemmer: ISRI Stemmer,” in 2019 IEEE 5th International Conference on Wireless and Telematics (ICWT), pp. 1–4, 2019.

[8] A. Al-Hasan, “Building a Sentiment Lexicon for the Palestinian Dialect,” Islamic Univercity, 2016.

[9] R. Piryani, D. Madhavi and V. K. Singh, “Analytical mapping of opinion mining and sentiment analysis re-search during 2000–2015”, Information Processing & Management, 53(1), 122-150, 2017.

[10] I. O. Hussien, K. Dashtipour, and A. Hussain, “Compar-ison of sentiment analysis approaches using modern Arabic and Sudanese Dialect,” in International Confer-ence on Brain Inspired Cognitive Systems, pp. 615–624, 2018.

[11] M. Jarrar, N. Habash, D. F. Akra, and N. Zalmout, “Building a corpus for Palestinian Arabic: a preliminary study,” 2014.

[12] M.A. Ibrahim and Salim N., Sentiment Analysis of Ara-bic Tweets: With Special Reference Restaurant Tweets. Int. J. Comput. Sci.Trends Technol., 4, 173–179. 2013.

[13] M. Al-Ayyoub, A. Nuseir, G. Kanaan and R. Al-Shalabi, Hierarchical Classifiers for Multi-Way Sentiment Anal-ysis of Arabic Reviews. Int. J. Adv. Comput. Sci. Appl., 7, 531–539, 2016.

[14] M.E.M. Abo, N. Idris, R. Mahmud, A. Qazi, I.A.T. Ha-shem, J.Z. Maitama, U. Naseem, S.K. Khan and S. Yang, A multi-criteria approach for arabic dialect sen-timent analysis for online reviews: Exploiting optimal machine learning algorithm selection. Sustainability, 13, 10018, 2021.

[15] J.O. Atoum and M. Nouman, Sentiment Analysis of Arabic Jordanian Dialect Tweets. Int. J. Adv. Comput. Sci., 10, 256–262, Appl. 2019.

[16] H. Najadat, A. Al-Abdi, and Y. Sayaheen, “Model-based sentiment analysis of customer satisfaction for the Jordanian telecommunication companies,” in 2018 9th International Conference on Information and Communication Systems (ICICS), pp. 233–237, 2018.

[17] I.S.A.M. Heamida, E.S.A.E. Ahmed, M.N.E. Mohamed and A.A.A.A. Salih, Applying Sentiment Analysis on Arabic comments in Sudanese Dialect. Int. J. Comput. Sci. Trends Technol., 8, 2020.

[18] R. Ismail, M. Omer, M. Tabir, N. Mahadi, and I. Amin, “Sentiment analysis for arabic dialect using supervised learning,” in 2018 International Conference on Com-puter, Control, Electrical, and Electronics Engineering (ICCCEEE), pp. 1–6, 2018.

[19] Z. Benmounah, A. Boulesnane, A. Fadheli, and M. Khial, “Sentiment Analysis on Algerian Dialect with Transformers”. Applied Sciences, 13(20), 11157, 2023.

[20] A. Elnagar, S. Yagi, A. B. Nassif, I. Shahin and S. A. Salloum, “Sentiment analysis in dialectal Arabic: a systematic review”. Advanced Machine Learning Technologies and Applications: Proceedings of AMLTA 2021, 407-417, 2021.

[21] A. Al-Thubaity, Q. Alqahtani, and A. Aljandal, “Senti-ment lexicon for sentiment analysis of Saudi dialect tweets,” Procedia Comput. Sci., vol. 142, pp. 301–307, 2018.

[22] N. Al-Twairesh, H. Al-Khalifa and A. Al-Salman, AraSenTi: Large-scale Twitter-specific Arabic sentiment lexicons. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 697-705), August, 2016.

[23] A. Assiri, A. Emam, and H. Al-Dossari, “Saudi twitter corpus for sentiment analysis,” World Acad. Sci. Eng. Technol. Int. J. Comput. Electr. Autom. Control Inf. Eng., vol. 10, no. 2, pp. 272–275, 2016.

[24] H. Mulki, H. Haddad, C. B. Ali and I. Babaoğlu, Tuni-sian dialect sentiment analysis: a natural language pro-cessing-based approach. Computación y Sistemas, 22(4), 1223-1232, (2018).

[25] S. R. El-Beltagy, “NileULex: a phrase and word level sentiment lexicon for Egyptian and modern standard Arabic,” in Proceedings of the Tenth International Con-ference on Language Resources and Evaluation (LREC’16), pp. 2900–2905, 2016.

[26] H. ElSahar and S. R. El-Beltagy, “A fully automated approach for arabic slang lexicon extraction from mi-croblogs,” in International conference on intelligent text processing and computational linguistics, pp. 79–91, 2014.

[27] S. M. Mohammad and P. D. Turney, “Crowdsourcing a word–emotion association lexicon,” Comput. Intell., vol. 29, no. 3, pp. 436–465, 2013.

[28] M. Youssef and S. R. El-Beltagy, “MoArLex: An Arabic sentiment lexicon built through automatic lexicon ex-pansion,” Procedia Comput. Sci., vol. 142, pp. 94–103, 2018.

[29] H. K. Aldayel and A. M. Azmi, “Arabic tweets senti-ment analysis–a hybrid scheme,” J. Inf. Sci., vol. 42, no. 6, pp. 782–797, 2016.

[30] R. Baly, A. Khaddaj, H. Hajj, W. El-Hajj, and K. B. Shaban, “Arsentd-lev: A multi-topic corpus for target-based sentiment analysis in arabic levantine tweets,” arXiv Prepr. arXiv1906.01830, 2019.

[31] “Assem’s Arabic Light Stemmer (BETA).” [Online]. Available: https://www.arabicstemmer.com/. [Ac-cessed: 03-Oct-2020].

[32] Python Developers, “Stemming and Lemmatization with Python NLTK (Arabic ISRI),” 2018. [Online]. Available: http://text-processing.com/demo/stem/. [Ac-cessed: 05-Sep-2020].

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

Mohammed KH. Zoroub holds a master degree in information technology from Islamic University of Gaza, Palestine. He is specialized in data science especially text mining and he published many research papers in this field.

Ashraf Y. Maghari is an associate professor of computer science at Islamic University of Gaza. He holds a PhD in computer vision and image processing from USM university, Malaysia. He has extensive research experience in computer science fields such as data/text mining, image processing, computer vision and deep learning.

Abdelkareem M. Alashqar received his PhD in Information Systems from Mansoura University, Egypt in 2017. He is an assistant professor at the department of software development, Islamic University of Gaza, Palestine. His research and teaching interests include software engineering, software quality evaluation, natural language processing and machine/deep learning.