



An automatic non-English sentiment lexicon builder using unannotated corpus

Mohammed Kaity¹ · Vimala Balakrishnan¹ 

Published online: 23 January 2019

© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract

Sentiment lexicons in the English language are widely accessible while in many other languages, these resources are extremely deficient. Current techniques and methods for sentiment analysis focus mainly on the English language, whereas other languages are neglected due to lack of resources. In order to overcome challenges faced in building non-English lexicons, we propose a language-independent method that automatically builds non-English sentiment lexicons based on currently available English lexicons with an unannotated corpus from the target language. The proposed method will automatically recognize and extract new polarity words from the unannotated corpus based on the initial seed lexicons that are developed by translating three reliable English lexicons. The experimental results from the test datasets confirmed that a developed non-English sentiment lexicon could significantly enhance the performance of non-English sentiment classifications, compared with other methods and lexicons. The developed lexicon in the Arabic language outperformed other commonly used methods for developing non-English lexicons, with an 0.74 *F* measure. The adopted approach in this study was proven to be language independent and can be implemented in other languages as well. This paper also contributes to understanding the approaches to developing sentiment resources.

Keywords Sentiment analysis · Natural language processing · Text analysis · Sentiment lexicon · Building resources

1 Introduction

Over the past few years, sentiment analysis which extracts embedded opinions found in a given dataset has received increasing interest, mostly due to the emergence and popularity of online social platforms (OSN). Previous studies have analysed

✉ Vimala Balakrishnan
vim.balakrishnan@gmail.com

¹ Department of Information System, Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur, Malaysia

opinions and views regarding products, services, news, social and political events, etc. [1], and thus providing insights into what people are thinking and feeling. The results from these analyses can be used in multiple applications such as propagation of hate speech [2].

Many sentiment analysis techniques exist, and one of the most commonly used technique uses the sentiment lexicon. A sentiment lexicon is described as a list of opinion or opinionated words and phrases based on their sentiment categories or orientations [3–5]. In the absence of an adequate training dataset, the lexicon-based approach is proven to be more appropriate than the machine learning approach [6]. Moreover, sentiment lexicons have been shown to perform better in short texts, such as social media texts [7]. The technique is also suitable for real-time opinion classification due to their relatively low requirements on computations [6, 8]. Furthermore, these sentiment lexicons can be employed for unsupervised [3, 6] and supervised classifications [5, 9] for a given text.

Sentiment lexicons are primarily available for the English language and completely limited or not available for other languages [10]. Interestingly, although English is recognized as the most commonly used language globally, the percentage of Internet users who communicate using English is less than 27%.¹ This means that it is dire if not an urgent need to create resources and tools for subjectivity purposes and sentiment analysis in non-English languages [11, 12]. Some researchers have attempted to build sentiment lexicons for non-English texts; however, they are not comparable to those used in the English language, as they are often incomplete or domain specific [11, 13]. An exciting and motivating factor towards creating and producing resources for non-English languages is supported by the fact that many organizations and enterprises recognize and appreciate the value and necessity to understand user feedback and associated trends, thereby gaining a competitive advantage regardless of language or demographics [14, 15]. Also, it is incredibly time-consuming and expensive to create sentiment lexicons for many languages manually [3, 16].

The extant of the literature has used translating methods to translate English lexicons to specific target languages to build non-English sentiment lexicons [13, 17, 18]; however, this results in the loss of context [19]. Others have used lexical language resources containing words with synonyms and antonyms, such as translated copies of SentiWordNet (SWN) as reported by [20, 21], where they built lexicons by analysing semantic relations between the words. This technique, however, is also limited as most non-English languages lack linguistic resources. Annotated corpus is another way to construct sentiment lexicons, which can be accomplished either using the statistical or the semantic relations method. Statistical techniques use large corpora with statistical equations to obtain the polarity of words, whilst semantic relations methods use semantic relations between the words in a large corpus to extract a sentiment lexicon [22]. The construction of sentiment lexicons by analysing the corpus requires a substantial volume of the corpus to enable an acceptable level of accuracy to be achieved. Moreover, in some instances, the annotated corpus requires additional data annotation and human effort before an analysis can commence [23, 24].

¹ <http://www.internetworldstats.com/stats7.htm>.

1.1 Contribution

The current study proposes an automatic language-independent, novel method for building non-English sentiment lexicons in order to address the gaps stated above. The proposed method uses existing English lexicons with unannotated target language corpus to identify the sentiment of the given document or word. The work will depend on the intuitive rule, whereby related words can determine the polarity of a sentiment word, and this rule will also hold true when determining the polarity of a document. The main contributions of the study are as follows:

- A novel framework incorporating two available resources (i.e. seed lexicons and unannotated corpus) is developed to build and adapt sentiment lexicons for non-English languages;
- The development of an automated method to recognize new polarity words in the unannotated corpus was determined by calculating the seed polarity values to predict the overall sentiment orientation of the candidate word (i.e. Eq. 1); and
- The construction of a curated Arabic sentiment lexicon based on the proposed method.

The remainder of this paper is structured accordingly. Section 2 describes the related work on lexicon generation methods for non-English languages. Section 3 provides an overview of the proposed automated method for building and expanding sentiment lexicons. Moreover, it presents the implementation of the method using real data and the evaluation, which is followed by the conclusion and suggestions for future work in Sect. 4.

2 Related work

2.1 Sentiment lexicon generation methods

Methods for generating opinion lexicons for the non-English languages range from being fully manual, semiautomatic, to restricted automatic methods [21]. Three common methods in generating these lexicons are: dictionary-based, corpus-based and human-based.

2.1.1 Dictionary-based method

The dictionary-based method relies on existing sentiment lexicons to build target language lexicons by translating, transferring learning or semantic relations [25]. Moreover, the rapid development of machine translation tools has enabled researchers to translate many English sentiment lexicons into other languages [17, 18]. Yao et al. [26] conducted one of the first studies in building sentiment lexicons by translation. They used a bilingual dictionary to determine the sentiment orientation of the Chinese words in order to generate a Chinese sentiment lexicon. The translation

technique was also used in Mihalcea et al. [27] to build a Romanian sentiment lexicon. On the other hand, Steinberger et al. [13] suggested a semiautomatic method to build sentiment lexicons by proposing a triangulation translating method between three languages, whereby two languages (English and Spanish) were used as sources, and the result of the translation was the third language. The authors found the triangulation technique to outperform the abstract translation in terms of building sentiment lexicons.

Other studies that have incorporated translation include the work of [18] who translated the English NRC Word Emotion Association Lexicon [28] to a French lexicon, and [29] in which the AFINN English sentiment lexicon used in [30] was translated to Norwegian language using Google translate.² Similarly, Basile and Nissim [31] used SentiWordNet and MultiWordNet to translate the sentiment orientation of English words to an Italian sentiment lexicon, whilst Perez-Rosas et al. [12] used OpinionFinder [32] along with SentiWordNet to transfer the sentiment scores from English to Spanish.

Though translation is a quick way to generate lexicons, it carries the risk of losing context and polarity associated with the words [25], which results in an inaccurate development of sentiment lexicons. Moreover, translation does not work on slangs and abbreviations that are commonly found on social networking sites [4].

2.1.2 Corpus-based method

The corpus-based method extracts the polarity words from a large volume of the corpus through the use of statistical or semantic relation methods [22]. Statistical-based methods use known polarity words (seeds) to exploit the co-occurrence of patterns to extract new sentiment words from a large corpus [24]. For example, Remus et al. [33] collected product reviews containing 5100 positive reviews and 5100 negative reviews and determined the polarity of the words using pointwise mutual information (PMI). Other studies that have used similar PMI include authors in [34] who created a Hindi sentiment lexicon, and [35] who generated an Arabic lexicon.

On the other hand, Elhawary and Elfeky [36] proposed a similarity graph to cluster all words/phrases to develop an Arabic sentiment lexicon. Their hypothesis was if two words have an edge, they are deemed to be similar; the similar words either have the same sentiment polarity or the same meaning. The random walk technique was used to build a Polish sentiment lexicon using 3222 web documents in [37], whereas emoticons were utilized to extract polarity words in a microblog for a Chinese lexicon in [38].

2.1.3 Human-based method

The human-based method relies on encouraging people to answer questions or solve puzzles to construct the sentiment lexicons from the answers [39]. Sentiment lexicons built by humans are usually more accurate than others [39]; however, the production of these lexicons is time-consuming, requires a lot of

² <https://translate.google.com>.

resources and is costly [28]. Nevertheless, studies that have engaged human experts to manually annotate data are many. For example, Abdul-Mageed et al. [40] proposed a supervised machine learning system called SAMAR, which analyses the Arabic subjectivity and sentiments in both Modern Standard Arabic and Arabic dialects by manually building a sentiment lexicon consisting of 3982 adjectives labelled as positive, negative or neutral.

One of the techniques gaining popularity in human-based approach is online games, whereby games are used to solicit 'experts' view. For instance, Lafourcade et al. [41] produced an online game with a purpose (GWAP) that requires the users to determine the polarity of the presented terms and words using emoticons whereas a language-independent crowdsourcing game called *Tower of Babel* that determines sentiments in Korean language was accomplished in [39]. Authors in [19] developed Sentiment Quiz and invited players from different countries to evaluate the given words. This resulted in more than 3500 users evaluating approximately 325,000 words in different languages. The authors, however, did not provide any information about their evaluations.

2.2 Limitations of current methods

All the three-mentioned techniques, that is, dictionary-based, corpus-based and human-based, have limitations. In short, the sentiment orientation of words and the sentiment lexicons built for dictionary-based approach tends to suit general domain lexicons, and thus may be inaccurate when used with specific domains. Moreover, sentiment lexicons do not contain many words or abbreviations commonly used on social networking sites, such as Twitter and Facebook. Therefore, the approach is not able to handle different dialects and informal/slang words [4]. Additionally, when machine translation is used errors may arise due to cultural differences about the sentiment orientations of words [12, 27, 42].

The corpus-based approach on the other hand lacks data pre-processing tools supporting many other languages, and thus making it difficult and complex to rely on the corpus to build lexicons. In addition, the generated lexicons cannot be relied upon to analyse other domains as they tend to be domain specific. The approach is also heavily dependent on annotated corpus [23] that requires manual data annotation. Finally, sentiment lexicons built by humans are usually more accurate than dictionary- and corpus-based approaches [39]; however, it is time-consuming and costly [28].

To overcome these challenges, the present study aims to build an automated language-independent non-English lexicon using unannotated corpus and English sentiment lexicons.

2.3 Arabic sentiment analysis

In this paper, the emphasis will be on the Arabic language sentiments. Consequently, this subsection describes some works of Arabic sentiment resources. Compared with the

available resources for the sentiment analysis of English and other languages, the Arabic language is severely under-resourced. With great interest in Arabic sentiment analysis in recent times [43], little research has been published compared to other languages. Some of these studies are based on the machine learning methods where researchers are labeling a corpus manually to positive or negative in order to get enough training data [44]. The vast majority of these corpora and resources are not available to the public [45]. On the other hand, some studies derive from the lexicon-based methods. Table 1 summarizes the main works in the building of Arabic sentiment lexicons.

In addition to the problems and limitations mentioned in Sect. 2.2, most of the Arabic lexicons are very noisy, since there is no part-of-speech (POS) and they are not available to the public [48]. The majority of lexicons were built for the Modern Standard Arabic language, which makes it appear weak when used with Arabic dialects [45]. This is because machine translation generates lexicon in the standard Arabic language.

3 Methodology

This section presents the proposed methodology for automatically constructing the non-English sentiment lexicons. A corpus-based method is proposed to discover new polarity words based on the following two resources: a target language corpus and an English seed sentiment lexicon. The seed sentiment lexicon is utilized to specify new sentiment words in the target language corpus. Figure 1 illustrates the process consisting of four phases: (1) seed lexicon preparation, (2) corpus collection and pre-processing, (3) candidate words extraction and (4) determination of the sentiment orientation for the candidate words. The steps involved in the four phases are presented in the following subsections.

3.1 Preparation of resources

3.1.1 Seed lexicon preparation

One of the main resources used in this study was the English sentiment lexicons, constructed by translating them to the target language using Google³ machine translation tools. The first step is to clean the translated lexicons by removing any duplicate and un-translated words. At this stage, more than one lexicon are used and combined to increase the number of words in the seed lexicon. If initial or sufficient seeds are available in the target language, then the automatic translation step will not take place.

³ <https://translate.google.com>.

Table 1 The main works in the building of Arabic sentiment lexicons

References	Data sources	Data sources references	Lexicon generating method	Domain	Number of entries	Evaluation method	Precision	Recall	F measure
Al-Twaresh et al. [35]	Opinion Lexicon MPQA	[46] [47]	Dictionary-based	Twitter	N/A	N/A	78.5	78.1	76.3
Al-Moslmi et al. [48]	MPQA	[47]	Dictionary-based and human-based	General	13,760	KNN SVM Others	N/A	N/A	77.97 82.07
El-Halees [49]	SentiStrength	[50]	Dictionary-based	Education, politics, sports	8793	LB	81.3	81.7	82.7
Hassan et al. [20]	WordNet Arabic WordNet General Inquirer	[51] [52] [53]	Dictionary-based	General	N/A	SO-PMI	83 93	N/A	N/A
Mahyoub et al. [54]	Arabic WordNet	[52]	Dictionary-based	General	7576	NB	94	91	N/A
Badaro et al. [55]	SAMA SentiWordNet Arabic WordNet	[56] [51, 57] [52]	Dictionary-based	General	28,812	SVM	58.3	95.1	72.3
Abdul-Mageed and Diab [58]	SIFAAT and HUDA SentiWordNet General Inquirer	[51, 59] [53]	Dictionary-based and human-based	General	224,564	LB	N/A	N/A	N/A
Elhawary and Elfeky [36]	large web corpus from the internet	N/A	Corpus-based	Business	1600	N/A	75–88	60–80	N/A
Eskander and Ram-bow [60]	AraMorph SentiWordNet	[61] [51]	Dictionary-based	General	35,000	SVM	67	66.6	68.6
Al-Twaresh et al. [35]	Tweets from Twitter	N/A	Human-based	Twitter	N/A	N/A	90.1	89.2	89.5
Al-Subaithin et al. [62]	Qaym. com players	N/A	Human-based	Business	N/A	N/A	N/A	N/A	N/A

SVM support vector machines, NB Naive Bayes, LB lexicon-based, LP label propagation test, KNN k-nearest neighbour

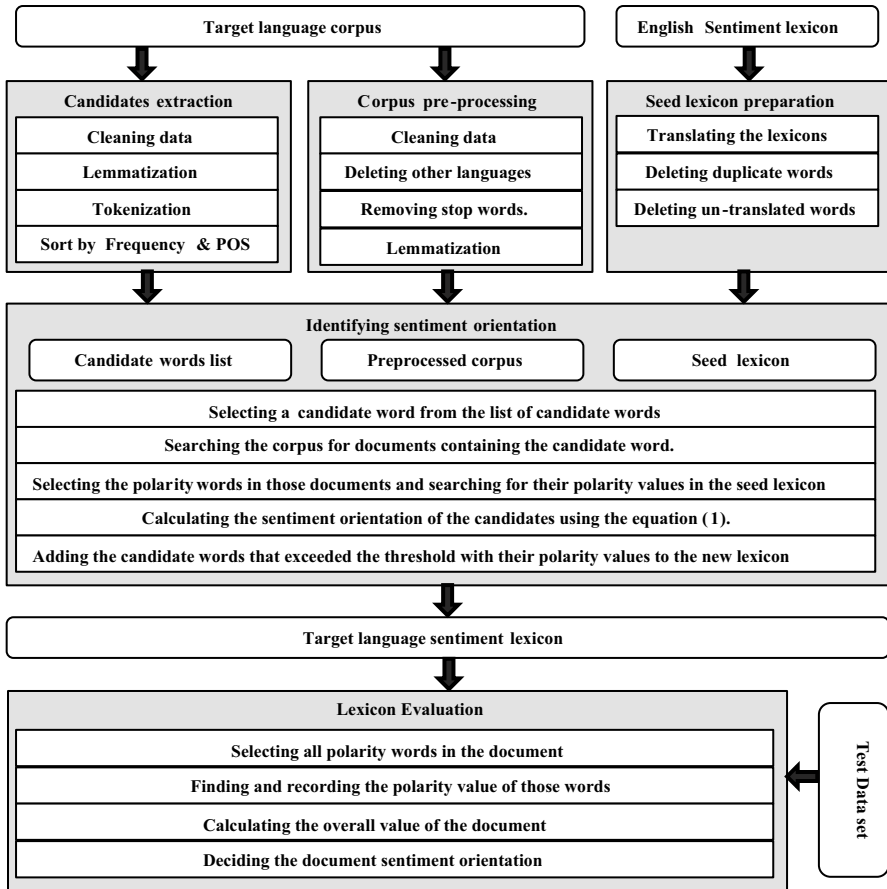


Fig. 1 The proposed method phases

3.1.2 Unannotated corpus preparation

An unannotated corpus was constructed using Facebook API⁴ due to the absence of annotated corpus for many languages [16]. Two corpus groups were established: one for the construction of the lexicon, and the other to test the accuracy of the classification. The pre-processing step involved data cleansing in which any comments that contained ‘links’ or ‘symbols’ were deleted, followed removing words and characters other than the words in the target language. ‘Common’ words (i.e. stop words), such as ‘he (هو)’, ‘you (أنت)’, ‘we (نحن)’ and ‘she (هي)’, were also removed. Finally, lemmatization (i.e. splitting off prefixes and affixes of the words) was performed to convert the words to their roots or dictionary

⁴ <https://developers.facebook.com>.

form [63]. For example, the sentence “the student’s books are different sizes” will be lemmatized in the following way;

the => the,
student’s => student,
books => book,
are => be,
different => differ,
sizes => size.

3.1.3 Candidate words extraction

The list of candidate words was prepared in several steps: first, the tokenization process that divides the sentences into individual words was carried out. For example, the sentence:

“the students go to school”
will become: “the”, “students”, “go”, “to”, “school”.

Next, the words were converted to their root or dictionary forms by splitting off the prefixes and affixes pertaining to the words.

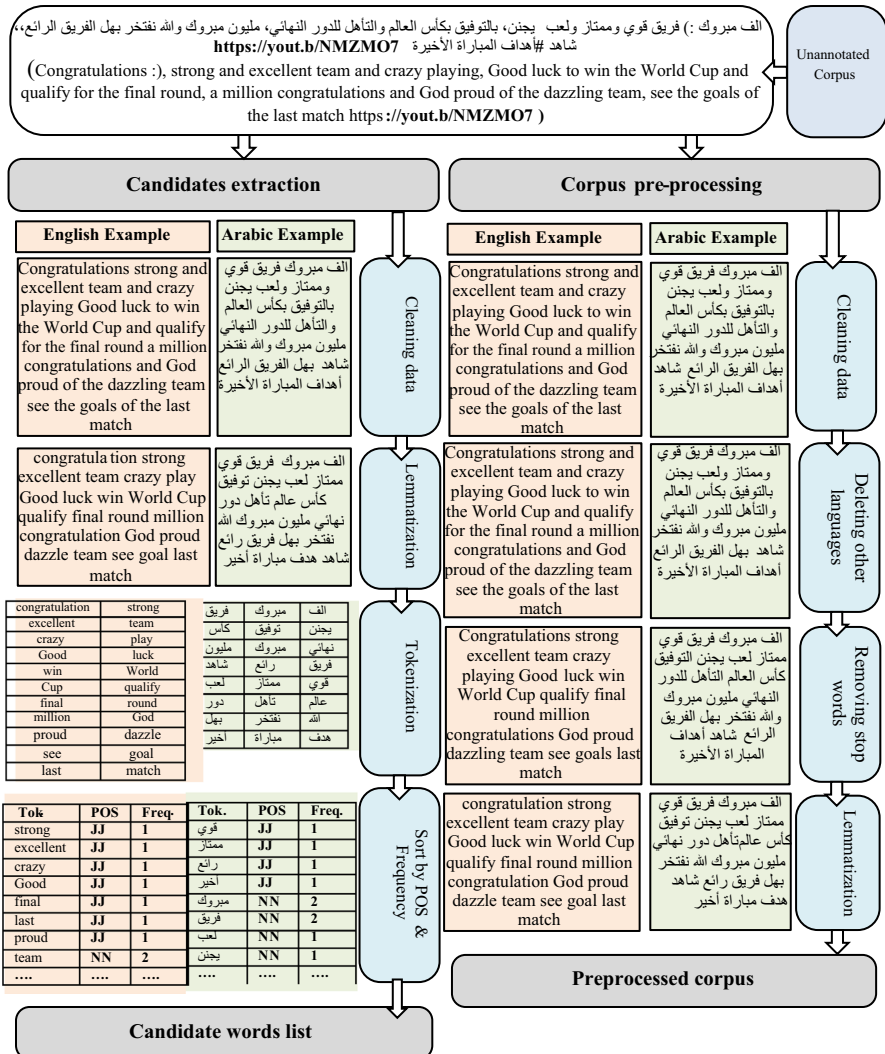
Several filters were applied to refine the lemmatized tokens. The first filter sorts the words alphabetically and removes words that are not contained within the target language, including any symbols or URLs. Additionally, any repeated or unusual words are also removed as these words are often misspelled or meaningless. The part-of-speech tags (POS) were then added to each candidate words in the list. Notably, adjectives and adverbs seem more likely to contain sentiment as shown in the previous studies [64, 65], unlike verbs and nouns [6]. Therefore, the present study gives priority to adjectives and adverbs by positioning them first during the candidate list sorting compared to verbs and nouns. Figure 2 shows an example of how the text pre-processing steps are accomplished to prepare the corpus and the candidate words list.

3.2 Sentiment orientation identification

The sentiment orientation of the candidate words was identified using the seed lexicon and pre-processed corpus, based on the relationships between the previously known polarity words (seeds) and the “new” words (candidates).

Figure 3 outlines the five steps which are performed to specify the candidate words’ sentiment orientation and are described as follows:

1. A new candidate word is selected from the candidate word list that was prepared in the previous phase (Sect. 3.1.3). Next,



Tok. = Token, Freq.= Frequency, POS=Part-of-Speech, JJ= Adjective, NN= Noun

Fig. 2 Resource preparation example

- The candidate word is used to search the corpus for any documents that contain the candidate word.
- The seed lexicon is used to specify the polarity words in those documents.
- Next, polarity values (*SWP*) are collected for the known seeds counting the number of words (*N*) and documents (*D*) and calculating the sentiment orientation of the candidate words (*P*) using Eq. (1).
- The final polarity class is calculated by Eq. (2), with (*T*) as the threshold value.

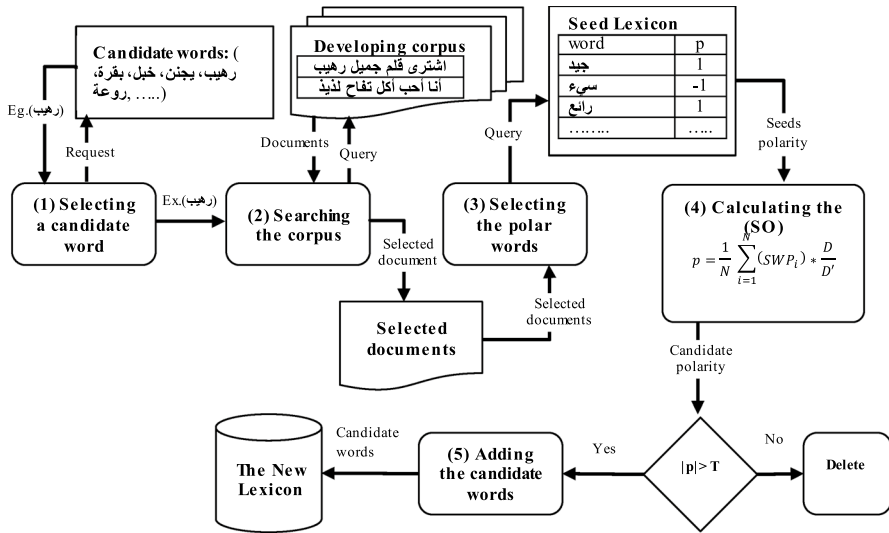


Fig. 3 The framework steps of building sentiment lexicons for non-English languages

The polarity equation (i.e. Eq. 1) is formulated based on the principle that a negative word should occur more frequently alongside the negative seed words and thus will obtain a negative score, whereas a positive word will occur most often in the vicinity of positive seed words, thus obtaining a positive score [66]. Equation (1) therefore depends on seed polarity values to predict the overall sentiment orientation of the candidate word, accomplished by aggregating the values of seed words in the documents and dividing by the total number of seed words.

$$\text{Candidate polarity } (w) = \frac{\sum \text{Polarity of nearby polar words}}{\sum \text{Nearby polar words}} * \text{The number of documents}$$

$$P(w) = \left(\frac{1}{N} \sum_{i=1}^N (SWP_i) \right) * \frac{D}{D'} \tag{1}$$

$$C(w) = \begin{cases} \text{Positive if } P(w) \geq (+T) \\ \text{Negative if } P(w) \leq (-T) \\ \text{Neutral else} \end{cases} \tag{2}$$

In Eq. (2) [4], (w) is a candidate word and (P) is the candidate word polarity value calculated by collecting the polarity values of the seed words (SWP) found in the same document, and divided by the number of seed words (N) obtained. Next, the result is multiplied by the number of selected documents (D) and divided by the total number of the documents (D'), and thus ensuring a minimal noise due to any

misspellings. The higher the frequency of a word in multiple documents, the more likely it is a polarity word. In some cases, however, the candidate words are only repeated numerous times in a single document (i.e. single author), suggesting that the occurrence of the word is likely to be misspelled.

For example, assume the candidate word is ‘lawl’, and the seed lexicon = [(good, 1); (happy, 1); (success, 1); (mistake, -1); (win, 1)], where positive = 1 and negative = -1, and the three selected documents are as follows:

“Lawl! That’s *good* news!”,
 “My mom will be *happy* because of my *success*! lawl!!”,
 “Lawl... despite their *mistakes*, they will *win*”, from a corpus contains four documents.

Based on the given seed lexicon, the seed words polarity (SWP) = (1, 1, 1, -1, 1), the count of seed words (N) = 5, the count of documents (D) = 3, and the total number of documents (D’) = 1000. The final result is calculated as follows:

$$p(\text{Lawl}) = \left(\frac{1}{N} \sum_{i=1}^N (\text{SWP}_i) \right) * \frac{D}{D'} = \left(\frac{1 + 1 + 1 + (-1) + 1}{5} \right) * \frac{3}{1000} = 0.0018$$

At first glance, the word can be construed as either positive [i.e. (P) is positive] or negative [i.e. (P) is negative]. However, doing so merely creates unnecessary noise in the lexicon due to the addition of weak polarity words. In this case, the value for threshold (T) is assumed to be both negative and positive. Moreover, if the positive value is greater or equal to the positive threshold (+T), then the word will be considered positive, and vice versa.

3.3 Experiments and evaluation

This section presents the experiments to evaluate the proposed method, with explanations provided for data collection, pre-processing and the evaluation of the lexicon.

3.3.1 Data collection and pre-processing

Two primary resources: (1) target language corpus and (2) seed lexicons are first collected. As mentioned in Sect. 1, the study focuses on the Arabic language and the choice of domain was media news.

Data were collected from Facebook, a social networking site with a global presence of more than 1 billion active users. These users subscribe to various services and often voice opinions and views through online posts, which are used as a communication tool to interact with their friends, relatives and user groups [67]. To be specific, the data were fetched from six Arabic Facebook official news sites (identities withheld for confidentiality purpose). An application programming interface (API) script was used to access and collect the data, resulting in a total of 20,816

Table 2 The steps of preparing candidate list with the number of tokens in each step

No	Steps	Number of tokens
1	Total token numbers	507,529 Tokens
2	Delete duplicate words	120,863 Tokens
3	Delete symbols and numbers	89,621 Tokens
4	Delete non-Arabic words	77,207 Tokens
5	Delete stop words	48,964 Tokens
6	Delete less frequent words	6821 Tokens
7	The candidate list	6821 Tokens

posts (i.e. equivalent to 507,529 tokens) during the period of 3 August 2017 to 23 August 2017.

The unannotated corpus was then cleaned by deleting posts that contained URL links and other symbols, as these are often associated with spam, advertisements or irrelevant comments. Words and characters other than the Arabic language were also removed, followed by the removal of the Arabic ‘stop words’ that was done by comparing the corpus with the lists available on the Internet.⁵ For lemmatizing the corpus, FARASA [68], which is a quick and reliable Arabic text processing toolkit, was used to convert all the words to their roots, and their dictionary forms. Candidate words were then extracted, and the final corpus contained 10,219 documents. Of these, 90% (9219) of the documents were used as the training set, whereas the remaining 10% (1000) were used as the testing set. The total number of tokens was 507,529. Table 2 shows the steps adopted to clean the tokens to achieve an appropriate set of candidate words.

Three English lexicons readily available on the Internet were collected and used to prepare the seed lexicon, as shown in Tables 3, 4 and Fig. 4. The three lexicons were AFINN [29], MPQA (Multi-Perspective Question Answering) [47] and Bing Liu’s Opinion Lexicon [46]. Google translate⁶ tool was used to translate the English lexicons to Arabic. Due to the nature of the generic translation, it is common for some synonyms to be translated to the same word, for example, the words *wonderful*, *terrific*, *marvellous*, *gorgeous* and *fabulous*, translated to the most frequently used synonym in the Arabic language is “رائع”. Therefore, in our proposed method, we collected sentiment words through the analysis of the corpus, as the corpus contains the words that people actually use on the internet.

This phase generates three outputs: (1) seed lexicon, (2) pre-processed corpus and (3) candidate words.

⁵ <https://www.ranks.nl/stopwords/arabic>.

⁶ <https://translate.google.com>.

Table 3 The English lexicons that used to build the seed lexicon

Sentiment lexicon	Domain	Entry size	Polarity	Description
Bing Liu's Opinion Lexicon [46]	General domain	6385 words	-1, 0, 1	Opinion lexicon is a manually created lexicon, and it includes 4783 negative words and 2006 positive words
AFINN [29]	General Domain	3382	-1, 0, 1	The first version contained 1468 unique words and called AFINN-96. The newest version has 2477 different words
MPQA Subjectivity Lexicon [47]	General domain	7937	Strong or weak positive or negative	MPQA lexicon consist of 7937 words with their subjectivities (strong or weak), polarities and POS tags

Table 4 Pre-processing stages of the seed lexicons

Lexicon	Size	After translating	Un-translated words	Duplicate entries	Total
MPQA	7937	7652	285	2974	4678
AFINN	3382	3265	117	838	2427
Bing Liu (OL)	6385	5982	403	1426	4556

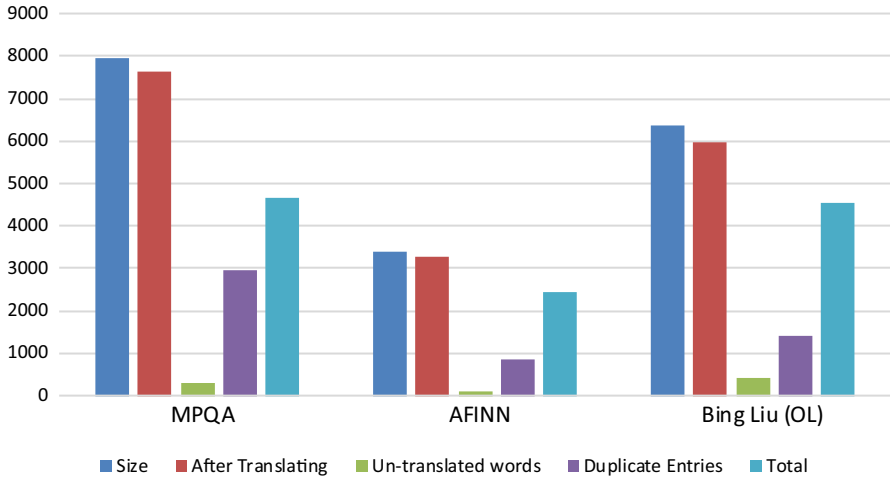


Fig. 4 Pre-processing stages of the seed lexicons

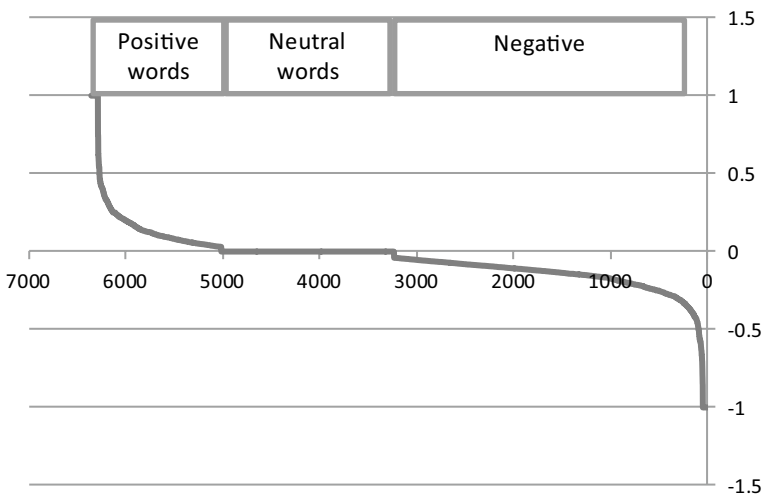


Fig. 5 The distribution of the polarity and neutral words

3.3.2 Experimental procedure

A candidate word from the candidate word list was first selected, which is then used as a parameter to query the pre-processed corpus of documents. The seed lexicon was then used to specify the polarity words and their polarity values in the documents. The total polarity value of the candidate word is calculated using Eq. (1), in Sect. 3.2. These steps were repeated for the rest of the words in the candidate word list.

The initial results showed many polarity words and neutral in various degrees. In the next step, the specific value for the threshold (T) value was set manually to determine the boundary between the polarity words and neutral words. The threshold was determined based on the absence or lack of words with a sentiment. The present study set the positive threshold ($+T$)=0.003, and the negative threshold ($-T$)=-0.004, with any values in between treated as neutral. Our calculation shows there were 1340 positive words, 3239 negative words and 1777 neutral words. Figure 5 illustrates the distribution of the polarity for these words, indicating the number of negative words to be far greater than positive words.

3.3.3 Evaluation

At total of 1000 posts from the lexicon were randomly selected and manually labelled as either positive, negative or neutral. The lexicon scoring method was adopted for classification purposes, i.e. in any document, if the number of positive words was higher than the number of negative words, then the document was classified as positive; else, it was classified as negative. The resulting lexicon was then used to train the classifier. A confusion matrix [69] was used with four indices: accuracy (A), precision (P), recall (R) and F measure (F), to measure the performance of the proposed method, based on the following equations:

$$A = \frac{TP + TN}{TP + TN + FP + FN}, \quad P = \frac{TP}{TP + FP}, \quad R = \frac{TP}{TP + FN}, \quad F = 2 \cdot \frac{P \cdot R}{P + R} \quad (3)$$

where TN (true negatives) are instances that are supposed to be negative and appeared as negative; TP (true positives) are instances that are supposed to be positive and appeared positive; FN (false negatives) are instances that are supposed to be positive but appeared negative; and FP (false positives) are instances that are supposed to be negative but appeared positive [70]. Table 5 presents an example of calculating the evaluation measures based on the confusion matrix. Seven cases were selected randomly from the test dataset which has been annotated manually to positive or negative (i.e. actual class). The predicted class indicates the classification using our method. For performance assessment, the results are aggregated as shown in Table 6, before applying Eq. (3).

The final result for this example is as follows:

$$\text{Accuracy} = \frac{2 + 3}{2 + 3 + 2 + 0} = 0.71$$

Table 5 An example of calculating the evaluation measures based on the confusion matrix

No	Arabic text (translation)	Actual class	Predicted class	Result
1	شعب عظيم، تحية للشعب العظيم شعب النضال، والكفاح، شعب الصمود، والتحدي، تحية لمصنع الرجال ومصدرها (A great people, greetings to the great people of struggle, struggle, people steadfastness, challenge, greetings to the factory of men and their source)	Positive	Positive	TP
2	اتفسحوا وافرحوا وانبتسطوا وسيبوا المستضعفين يقتلوا ويذبحوا حسبي الله ونعم الوكيل (Be satisfied, happy, and pleased, and let the vulnerable are killed and slaughtered. Allah suffices me, for He is the best disposer of affairs)	Negative	Positive	FP
3	وفي النهاية المجرمين يمزقون الاتفاقية ويهاجمون مسجد صلاح الدين ويستهدفون النساء والأطفال (In the end, criminals tear up the agreement and attack Salahuddin mosque and target women and children)	Negative	Negative	TN
4	يوم جميل ورهيب في حياة الشعب سيكون سببا في رفع الظلم عنه إن شاء الله (A beautiful and awesome day in the life of the people will be a reason to lift the injustice from them, God willing)	Positive	Positive	TP
5	انتم دعاة فتن لا يمكن بصدقكم الشعب الصادق النزوية (You are Advocates of sedition, so the fair and honest people cannot believe you)	Negative	Positive	FP
6	3 سنوات من القصف والقتل للمدنيين والكوليرا اكملت الحسبة حسينا الله ونعم الوكيل (3 years of bombing and killing of civilians and cholera completed them)	Negative	Negative	TN
7	هذا الخبر كلام فاضي ما يحتاج تحليل ولا تعب خلوكم في حالكم احسن. (This news is empty words, do not needs analysis and fatigue. it is better to be in your business.)	Negative	Negative	TN

TP true positives, *TN* true negatives, *FN* false negatives, *FP* false positives

$$\text{Precision} = \frac{2}{2+2} = 0.5$$

$$\text{Recall} = \frac{2}{2+0} = 1$$

$$F \text{ measure} = 2 \cdot \frac{0.5 * 1.0}{0.5 + 1.0} = 0.67$$

Because of the scarcity of Arabic sentiment lexicons and their lack of availability to the public, we compared our method with two lexicons, namely AraSenTi [35] and Arabic-NRC [71]. Moreover, we translated three English lexicons to compare the translation-based methods with our method. In Sect. 2, we mentioned that previous studies used the manual method of building some sentiment lexicons, but we could not find any available Arabic lexicon that was built manually to compare with our lexicon. A brief description of the lexicons is as follows:

Table 6 The confusion matrix of the example in Table 5

	Actual class	
	Positive	Negative
Predicted class		
Positive	TP=2	FP=2
Negative	FN=0	TN=3

TP true positives, *TN* true negatives, *FN* false negatives, *FP* false positives

- *T_MPQA* is the translated copy of the MPQA sentiment lexicon translated by Google translator.⁷
- *T_OL* is the translated copy of the Bing Liu's opinion lexicon translated by Google translator.
- *T_AFINN* is the translated copy of the AFINN sentiment lexicon translated by Google translator.
- *Hybrid 3* is the combination of the three translated lexicons *T_MPQA*, *T_OL*, and *T_AFINN*.
- *Unannotated Corpus-Based Sentiment Lexicon (UCBSL)*⁸ is the proposed sentiment lexicon developed by the method outlined in Sect. 3 in this study.
- *Hybrid 4* is the combination of (*Hybrid 3*) and the proposed lexicon (*UCBSL*).
- *AraSenTi (Arabic)* is the large-scale Arabic sentiment lexicon generated from a large dataset for social network sentiment analysis [35].
- *Arabic-NRC (sentiment lexicon)* The NRC emotion lexicon included emotional English words which were divided with their POS (adjectives, verbs, nouns and adverbs) and positive and negative sentiments [71]. The authors translated the Arabic version.⁹

Table 7 and Fig. 6 list the number of polarity entries in those sentiment lexicons.

3.3.3.1 Human evaluation The Arabic lexicon generated was manually checked and validated by five professional Arabic linguists. Copies of 200 randomly sampled cases of the lexicon words were provided to them, and then we asked them to identify positive and negative words. The results of the manual validations were compared with the polarity values generated automatically by our method. Average accuracy was obtained at 81%, which means that our method could lead to wrong inputs of 19%. The outcome indicating the proposed automatic non-English sentiment lexicon builder is capable of producing a good quality lexicon. Table 8 lists some polarity words from our lexicon list with their sentiment orientation and frequency.

⁷ <https://translate.google.com>.

⁸ Available online in: <https://github.com/mohkaity/UCBSL>.

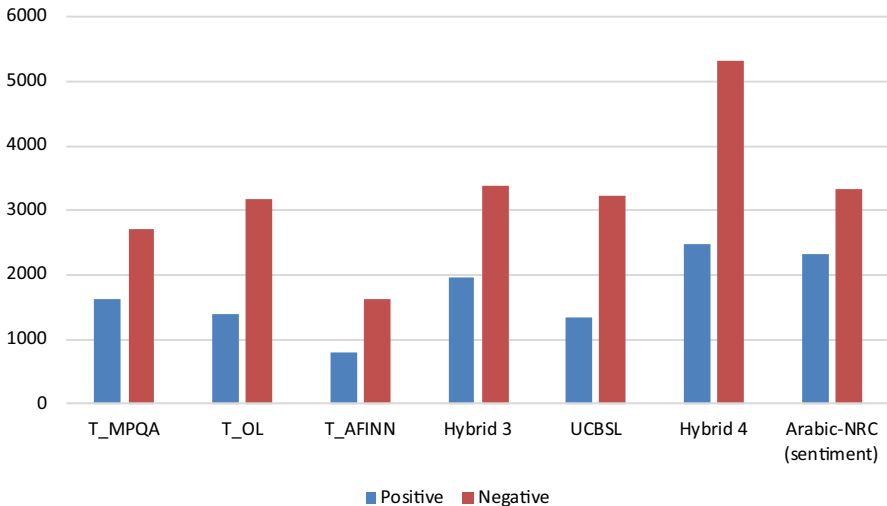
⁹ <http://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm>.

Table 7 Lists the numbers of positive and negative entries in examined lexicons

Lexicon	Building method	Positive	Negative	Neutral	Total
T_MPQA	Dictionary-based (translation)	1637	2716	325	4678
T_OL	Dictionary-based (translation)	1388	3168	–	4556
T_AFINN	Dictionary-based (translation)	806	1619	2	2427
Hybrid 3	T_MPQA + T_OL + T_AFINN	2223	3953	294	6470
UCBSL	Corpus-based	1340	3239	1777	6356
Hybrid 4	T_MPQA + T_OL + T_ AFINN + UCBSL	2478	5316	1916	9709
AraSenTi (Arabic)	Corpus-based	116,448	108,881	–	225,329
Arabic-NRC (sentiment)	Dictionary-based (translation)	2312	3325	–	5637

T_MPQA translated copy of the MPQA, *T_OL* translated copy of the Bing Liu's lexicon, *T_AFINN* translated copy of the AFINN, *Hybrid 3* combination of T_MPQA, T_OL, and T_AFINN, *UCBSL* unannotated corpus-based sentiment lexicon, *Hybrid 4* combination of (Hybrid 3) and (UCBSL), *AraSenTi* large-scale Arabic sentiment, *Arabic-NRC* the NRC emotion lexicon

Bolded values indicate results of the proposed model

**Fig. 6** The numbers of positive and negative entries in the examined lexicons (except AraSenTi)

4 Results and discussion

4.1 Experimental results

Table 9 and Fig. 7 show the performance results of the new lexicon compared with other sentiment lexicons discussed in Sect. 3.3.3. The results show that the new sentiment lexicon outperformed the other sentiment lexicons, achieving 0.74 *F* measure. The closest lexicon with regard to the *F* measure was the Hybrid 4, which attained an *F* measure of 0.69. The Hybrid 4 lexicon included the new lexicon and the three seed lexicons T_MPQA, T_OL and T_AFINN. However,

Table 8 Sample sentiment words from our lexicon

No.	Word (Arabic)	POS	Arabic transliteration	Meaning	Frequency	Polarity value	Class
1	قتل	Noun	<i>qutil</i>	Killing	6691	-0.22149	Negative
2	حرب	Noun	<i>harb</i>	War	5645	-0.18636	Negative
3	فتنة	Noun	<i>fitna</i>	Sedition	1645	-0.12158	Negative
4	مرتزق	Adjective	<i>murtaziq</i>	Mercenary	492	-0.26016	Negative
5	همجي	Adjective	<i>himji</i>	Barbaric	257	-0.26848	Negative
6	خير	Adjective	<i>khayr</i>	Good	3097	0.07362	Positive
7	عظيم	Adjective	<i>eazim</i>	Great	2253	0.08966	Positive
8	سلام	Noun	<i>salam</i>	Peace	2152	0.10037	Positive
9	سبحان	Adjective	<i>subhan</i>	Glory	1506	0.13413	Positive
10	جنة	Noun	<i>jana</i>	Paradise	1039	0.13763	Positive

Table 9 The performance results of our lexicon compared to a number of sentiment lexicons

Lexicon	Accuracy	Precision	Recall	<i>F</i> measure
T_MPQA	0.47	0.63	0.68	0.60
T_OL	0.49	0.62	0.67	0.62
T_AFINN	0.50	0.68	0.73	0.65
Hybrid 3	0.58	0.68	0.72	0.67
UCBSL	0.78	0.78	0.72	0.74
Hybrid 4	0.74	0.71	0.68	0.69
AraSenTi (Arabic)	0.59	0.61	0.63	0.57
Arabic-NRC (sentiment)	0.45	0.62	0.65	0.55

Bolded values indicate results of the proposed model

the translated lexicons did not achieve *F* measure exceeding 0.67 including NRC, which was the sentiment lexicon translated by its authors.

The AraSenTi sentiment lexicon consisted of many polarity words generated through translation and using PMI statistical equation; however, the *F* measure of this lexicon did not exceed 0.57. This means that the size of the lexicon is not always useful, and maybe more of an issue, given this is considered as a major challenge in performing sentiment analysis [72]. We expected Hybrid 4 to achieve a higher *F* measure as it contained both the new lexicon (UCBSL) and three translated lexicons (Hybrid 3); however, this was not the case. This is probably due to the translation process in Hybrid 3 that may have resulted in incorrect or inaccurate polarity words originating from the seed lexicons. For example, the word “terrible” is translated to the Arabic word “رهيب” in Hybrid 4, while Arab users generally use this word in informal languages to express a positive sentiment rather than negative. Therefore, the translation process in Hybrid 3 may have affected the performance of Hybrid 4. This is also observed in Fig. 7 where Hybrid 3 produced the lowest performance compared to UCBSL and Hybrid 4.

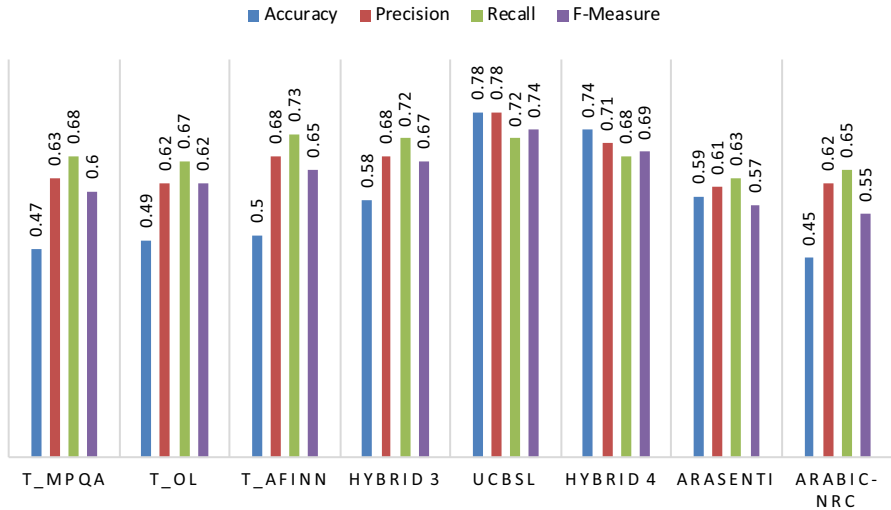


Fig. 7 The performance results of our lexicon compared to a number of sentiment lexicons

Table 10 The test results for both positive and negative classes

Lexicon	Positive			Negative		
	Precision	Recall	F measure	Precision	Recall	F measure
T_MPQA	0.35	0.74	0.47	0.90	0.62	0.73
T_OL	0.36	0.60	0.45	0.88	0.73	0.79
T_AFINN	0.44	0.87	0.58	0.92	0.59	0.72
Hybrid 3	0.47	0.78	0.58	0.89	0.66	0.76
UCBSL	0.73	0.51	0.60	0.83	0.92	0.87
Hybrid 4	0.60	0.48	0.53	0.81	0.87	0.84
AraSenTi (Arabic)	0.37	0.72	0.48	0.84	0.53	0.65
Arabic-NRC	0.34	0.85	0.49	0.90	0.44	0.60

Bolded values indicate results of the proposed model

The results were observed to differ at the class level, as depicted in Table 10. Across all the lexicons, the precision and *F* measure in the negative class showed much better results than in the positive class. For instance, in the new lexicon, the *F* measure in the negative class was 0.87 while it was only 0.60 in the positive class. This is probably due to the fact that the seed lexicon in the study contained more negative words than positive words.

The low recall values of translated lexicons (i.e. seed lexicons) in the negative class indicate that the classic lexicons clearly suffer from coverage problem and incorrect polarity values when used to classify social media data. On the other hand,

Table 11 The intersection of the words of the obtained lexicons with other lexicons

Lexicon	T_MPQA	T_OL	T_AFINN	Hybrid 3	UCBSL	Hybrid 4
Hybrid 3	0.83	0.82	0.49	0.100	0.21	0.39
UCBSL	0.21	0.16	0.19	0.21	0.100	0.79
Hybrid 4	0.30	0.28	0.18	0.39	0.79	0.100

Bolded values indicate results of the proposed model

our lexicon (UCBSL) suffers from a low recall of 0.51 in the positive class, probably due to the lack of positive words in the used corpus. We, however, increased the seed lexicons to our automatically generated lexicon to improve the recall as shown in Hybrid 4 (i.e. recall = 0.78).

Table 11 presents the intersection of the words of the new lexicons with existing lexicons. The table indicates that the UCBSL lexicon contains new entries not available in the translated lexicons, with results showing the rate of agreement between the UCBSL lexicon and Hybrid 3 is 21% (i.e. meaning UCBSL contains about 79% of new entries).

As shown in Table 9, the UCBSL lexicon outperformed the Arabic available lexicons built by dictionary and corpus-based methods. Due to the difficulty of building manual lexicons for each language and domain, our proposed method will facilitate the construction of new lexicons or expansion and adaptation of existing lexicons in a much effective and easy manner. As a result, the methodology, used in this study, supports those methods that use co-occurrence-based measures to find implicit relationships in unstructured data. Moreover, this study demonstrated that the related words could, in fact, determine the polarity of a sentiment word in the same context. An important criterion for the adoption of a 'new' word is based on the coverage or spread of the 'word' amongst users and repetition. The word is deemed entirely worthless if it is not commonly used amongst users or frequently repeated in documents. Furthermore, the methodology used for the experiments is language independent and can, therefore, be implemented for use in many other languages.

5 Conclusions

One of the major limitations in sentiment classification for non-English languages is that most of the existing annotated corpora are in English [16]. Many limitations and shortcomings reported in prior studies were addressed in the present study by using available resources and minimizing the amount of human effort in data labelling. This was accomplished by developing an automatic method for building non-English sentiment lexicons using two types of available and relatively cheap resources, namely the target language unannotated corpus and the English seed sentiment lexicons. The proposed method was evaluated using Arabic posts gathered from Arabic news media on Facebook. The evaluation results showed that the new Arabic lexicon

produced the highest accuracy with an F measure of 0.74, compared to translated lexicons and other Arabic lexicons.

The study is not without its limitations. For example, dealing with non-English languages further presents a number of difficulties and limitations such as the limited size of resources or their availability to the public [44]. Moreover, the deficiency and availability of pre-processing data tools for some languages continue to be a concern. This affected the study in the sense that limited Arabic sentiment lexicons were available for us to compare with the new lexicon. In addition, the performance of the proposed method may also have been affected due to the nature of the language as well, as the Arab people frequently write in multiple dialects, incurring numerous spelling and typographical errors [73]. In addition, diacritics (i.e. marks used to represent vowel sounds in Arabic script) are often used in formal Arabic communication [74]; hence, they were not analysed in the current study. Future studies could look into comparing diacritic Arabic with informal Arabic communication such as on social media.

The current study only evaluated the proposed method on a single language, and thus the assessment is limited. It would be interesting to examine the performance of the proposed method in other languages such as French or Dutch. Further, future studies could also explore other performance measures such as time complexity involved in generating the lexicons automatically. Finally, social media features such as emoticons and hashtags were not included in the present study, and this can be another avenue for future studies to explore.

References

1. Vilares D, Alonso MA, Gómez-Rodríguez C (2017) Supervised sentiment analysis in multilingual environments. *Inf Process Manag* 53(3):595–607
2. Williams ML, Burnap P (2015) Cyberhate on social media in the aftermath of Woolwich: a case study in computational criminology and big data. *Br J Criminol* 56(2):211–238
3. Bravo-Marquez F, Frank E, Pfahringer B (2016) Building a Twitter opinion lexicon from automatically-annotated tweets. *Knowl Based Syst* 108:65–78
4. Wu FZ, Huang YF, Song YQ, Liu SX (2016) Towards building a high-quality microblog-specific Chinese sentiment lexicon. *Decis Support Syst* 87:39–49
5. Kiritchenko S, Zhu X, Mohammad SM (2014) Sentiment analysis of short informal texts. *J Artif Intell Res* 50:723–762
6. Deng S, Sinha AP, Zhao H (2017) Adapting sentiment lexicons to domain-specific social media texts. *Decis Support Syst* 94:65–76
7. Bermingham A, Smeaton AF (2010) Classifying sentiment in microblogs: is brevity an advantage? In: *Proceedings of the 19th ACM International Conference on Information and Knowledge Management, ACM*
8. Chaovalit P, Zhou L (2005) Movie review mining: a comparison between supervised and unsupervised classification approaches. In: *Proceedings of the 38th Annual Hawaii International Conference on System Sciences, 2005. HICSS'05, IEEE*
9. Kouloumpis E, Wilson T, Moore JD (2011) Twitter sentiment analysis: the good the bad and the omg! *ICWSM 11(538–541):164*
10. Wu S-J, Chiang R-D, Ji Z-H (2017) Development of a Chinese opinion-mining system for application to Internet online forums. *J Supercomput* 73(7):2987–3001
11. Lo SL, Cambria E, Chiong R, Cornforth D (2017) Multilingual sentiment analysis: from formal to informal and scarce resource languages. *Artif Intell Rev* 48(4):499–527

12. Perez-Rosas V, Banea C, Mihalcea R (2012) Learning sentiment lexicons in Spanish. In: *Lrec 2012: Eighth International Conference on Language Resources and Evaluation*, 2012, pp 3077–3081
13. Steinberger J, Ebrahim M, Ehrmann M, Hurriyetoglu A, Kabadjov M, Lenkova P, Steinberger R, Tanev H, Vázquez S, Zavarella V (2012) Creating sentiment dictionaries via triangulation. *Decis Support Syst* 53(4):689–694
14. Liu B (2012) Sentiment analysis and opinion mining. *Synth Lect Hum Lang Technol* 5(1):1–167
15. Lo SL, Cambria E, Chiong R, Cornforth D (2016) A multilingual semi-supervised approach in deriving Singlish sentic patterns for polarity detection. *Knowl Based Syst* 105:236–247
16. Sun S, Luo C, Chen J (2017) A review of natural language processing techniques for opinion mining systems. *Inf Fusion* 36:10–25
17. Dashtipour K, Poria A, Hussain A, Cambria E, Hawalah AY, Gelbukh A, Zhou Q (2016) Multilingual sentiment analysis: state of the art and independent comparison of techniques. *Cogn Comput* 8(4):757–771
18. Abdaoui A, Azé J, Bringay S, Poncelet P (2017) Feel: a french expanded emotion lexicon. *Lang Resour Eval* 51(3):833–855
19. Scharl A, Sabou M, Gindl S, Rafelsberger W, Weichselbraun A (2012) Leveraging the wisdom of the crowds for the acquisition of multilingual language resources
20. Hassan A, Abu-Jbara A, Jha R, Radev D (2011) Identifying the semantic orientation of foreign words. In: *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Short Papers*, vol 2. Association for Computational Linguistics
21. Nusko B, Tahmasebi N, Mogren O (2016) Building a sentiment lexicon for swedish. In: *Digital Humanities 2016. From Digitization to Knowledge 2016: Resources and Methods for Semantic Processing of Digital Works/Texts*, Proceedings of the Workshop, 11 July 2016, Krakow, Poland. Linköping University Electronic Press
22. Kumar P, Jaiswal UC (2016) A comparative study on sentiment analysis and opinion mining. *Int J Eng Technol* 8(2):938–943
23. Pozzi FA, Fersini E, Messina E, Liu B (2017) Chapter 1: challenges of sentiment analysis in social networks: an overview. *sentiment analysis in social networks*. Morgan Kaufmann, Boston, pp 1–11
24. Zhang HL, Gan WY, Jiang B (2014) IEEE, machine learning and lexicon based methods for sentiment classification: a survey. In: *2014 11th Web Information System and Application Conference (WISA)*, 2014, pp 262–265
25. Denecke K (2008) Using sentiwordnet for multilingual sentiment analysis. In: *IEEE 24th International Conference on Data Engineering Workshop*, 2008. ICDEW 2008, IEEE
26. Yao J, Wu G, Liu J, Zheng Y (2006) Using bilingual lexicon to judge sentiment orientation of Chinese words. In: *The Sixth IEEE International Conference on Computer and Information Technology*, 2006. CIT'06, IEEE
27. Mihalcea R, Banea C, Wiebe JM (2007) Learning multilingual subjective language via cross-lingual projections
28. Mohammad SM, Turney PD (2013) Crowdsourcing a word–emotion association lexicon. *Comput Intell* 29(3):436–465
29. Nielsen FA (2011) A new ANEW: evaluation of a word list for sentiment analysis in microblogs. In: *1st Workshop on Making Sense of Microposts 2011: Big Things Come in Small Packages, #MSM 2011—Co-located with the 8th Extended Semantic Web Conference, ESWC 2011*. Heraklion, Crete
30. Hammer H, Bai A, Yazidi A, Engelstad P (2014) Building sentiment lexicons applying graph theory on information from three norwegian thesauruses. In: *Norsk Informatikkonferanse (NIK)*
31. Basile V, Nissim M (2013) Sentiment analysis on Italian tweets. In: *Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*
32. Wilson T, Hoffmann P, Somasundaran S, Kessler J, Wiebe J, Choi Y, Cardie C, Riloff E, Patwardhan S (2005) OpinionFinder: a system for subjectivity analysis. In: *Proceedings of HLT/EMNLP on Interactive Demonstrations*. Association for Computational Linguistics
33. Remus R, Quasthoff U, Heyer G (2010) SentiWS: a publicly available German-language resource for sentiment analysis. In: *LREC*
34. Jha V, Savitha R, Hebbar SS, Shenoy PD, Venugopal K (2015) HmSdS: Hindi multi-domain sentiment aware dictionary. In: *International Conference on Computing and Network Communications (CoCoNet)*, 2015, IEEE
35. Al-Twairesh N, Al-Khalifa H, Al-Salman A (2016) AraSenTi: large-scale twitter-specific Arabic sentiment lexicons. In: *Association for Computational Linguistics*, 2016, pp 697–705

36. Elhawary M, Elfeky M (2010) Mining Arabic business reviews. In: IEEE International Conference on Data Mining Workshops (ICDMW), 2010, IEEE
37. Haniewicz K, Kaczmarek M, Adamczyk M, Rutkowski W (2014) Polarity lexicon for the Polish language: design and extension with random walk algorithm. In: Swiatek J et al. (eds) International Conference on Systems Science, ICSS 2013, 2014. Springer, pp 173–182
38. Feng S, Song KS, Wang DL, Yu G (2015) A word-emoticon mutual reinforcement ranking model for building sentiment lexicon from massive collection of microblogs. *World Wide Web-Internet Web Inf Syst* 18(4):949–967
39. Hong Y, Kwak H, Baek Y, Moon S (2013) Tower of babel: a crowdsourcing game building sentiment lexicons for resource-scarce languages. In: 22nd International Conference on World Wide Web, WWW 2013, Rio de Janeiro
40. Abdul-Mageed M, Diab M, Kübler S (2014) SAMAR: subjectivity and sentiment analysis for Arabic social media. *Comput Speech Lang* 28(1):20–37
41. Lafourcade M, Joubert A, Le Brun N (2015) Collecting and evaluating lexical polarity with a game with a purpose. In: RANLP
42. Mohammad SM, Salameh M, Kiritchenko S (2016) How translation alters sentiment. *J Artif Intell Res* 55:95–130
43. Shboul BA, Al-Ayyoub M, Jararweh Y (2015) Multi-way sentiment classification of Arabic reviews. In: 2015 6th International Conference on Information and Communication Systems (ICICS)
44. Abdullah M, Hadzikadic M (2017) Sentiment analysis on Arabic Tweets: challenges to dissecting the language. In: International Conference on Social Computing and Social Media, 2017. Springer
45. Najar D, Mesfar S (2017) Opinion mining and sentiment analysis for Arabic on-line texts: application on the political domain. *Int J Speech Technol* 20:575–585
46. Hu M, Liu B (2004) Mining and summarizing customer reviews. In: Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM
47. Wilson T, Wiebe J, Hoffmann P (2005) Recognizing contextual polarity in phrase-level sentiment analysis. In: Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing, 2005. Association for Computational Linguistics
48. Al-Moslemi T, Albared M, Al-Shabi A, Omar N, Abdullah S (2018) Arabic senti-lexicon: constructing publicly available language resources for Arabic sentiment analysis. *J Inf Sci* 44(3):345–362
49. El-Halees A (2011) Arabic opinion mining using combined classification approach. In: The International Arab Conference on Information Technology, pp 10–13
50. Thelwall M, Buckley K, Paltoglou G, Cai D, Kappas A (2010) Sentiment strength detection in short informal text. *J Am Soc Inf Sci Technol* 61(12):2544–2558
51. Baccianella S, Esuli A, Sebastiani F (2010) SentiWordNet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In: LREC
52. Black W, Elkateb S, Rodriguez H, Alkhalifa M, Vossen P, Pease A, Fellbaum C (2006) Introducing the Arabic wordnet project. In: Proceedings of the Third International WordNet Conference
53. Stone PJ, Dunphy DC, Smith MS (1966) The general inquirer: a computer approach to content analysis. MIT Press, Oxford
54. Mahyoub FHH, Siddiqui MA, Dahab MY (2014) Building an Arabic sentiment lexicon using semi-supervised learning. *J King Saud Univ Comput Inf Sci* 26(4):417–424
55. Badaro G, Baly R, Hajj H, Habash N, El-Hajj W (2014) A large scale Arabic sentiment lexicon for Arabic opinion mining. *ANLP* 2014:165
56. Maamouri M, Graff D, Bouziri B, Krouna S, Bies A, Kulick S (2010) Standard Arabic morphological analyzer (SAMA) version 3.1. Linguistic Data Consortium, Catalog No.: LDC2010L01
57. Esuli A, Sebastiani F (2007) SentiWordNet: a high-coverage lexical resource for opinion mining. *Evaluation* 17:1–26
58. Abdul-Mageed M, Diab MT (2014) SANA: a large scale multi-genre, multi-dialect lexicon for Arabic subjectivity and sentiment analysis. In: LREC, 2014
59. Abdul-Mageed M, MT Diab (2011) Subjectivity and sentiment annotation of modern standard Arabic newswire. In: Proceedings of the 5th Linguistic Annotation Workshop, 2011. Association for Computational Linguistics
60. Eskander R, Rambow O (2015) SLSA: a sentiment lexicon for Standard Arabic. In: Conference on Empirical Methods in Natural Language Processing, EMNLP 2015. Association for Computational Linguistics (ACL)

61. Buckwalter T (2002) Buckwalter Arabic morphological analyzer version 2.0. Linguistic Data Consortium, University of Pennsylvania, 2002. LDC Catalog No.: LDC2004L02. 2004, ISBN 1-58563-324-0
62. Al-Subaihin AA, Al-Khalifa HS, Al-Salman AS (2011) A proposed sentiment analysis tool for modern arabic using human-based computing. In: Proceedings of the 13th International Conference on Information Integration and Web-Based Applications and Services, 2011, ACM
63. Abdul-Mageed M (2019) Modeling Arabic subjectivity and sentiment in lexical space. *Inf Process Manag* 56(2):291–307
64. Das SR, Chen MY (2007) Yahoo! for Amazon: sentiment extraction from small talk on the web. *Manag Sci* 53(9):1375–1388
65. Velikovich L, Blair-Goldensohn S, Hannan K, McDonald R (2010) The viability of web-derived polarity lexicons. In: 2010 Human Language Technologies Conference of the North American Chapter of the Association for Computational Linguistics, NAACL HLT 2010, Los Angeles, CA
66. Taboada M, Brooke J, Tofiloski M, Voll K, Stede M (2011) Lexicon-based methods for sentiment analysis. *Comput Linguist* 37(2):267–307
67. Davalos S, Merchant A, Rose GM, Lessley BJ, Teredesai AM (2015) ‘The good old days’: an examination of nostalgia in Facebook posts. *Int J Hum Comput Stud* 83:83–93
68. Abdelali A, Darwish K, Durrani N, Mubarak H (2016) Farasa: a fast and furious segmenter for Arabic. In: HLT-NAACL Demos, 2016
69. Powers D (2007) Evaluation: from precision, recall and fmeasure to roc, informedness, markedness and correlation. *J Mach Learn Technol* 2:37–63
70. Giachanou A, Crestani F (2016) Like it or not: a survey of twitter sentiment analysis methods. *ACM Comput Surv (CSUR)* 49(2):28
71. Mohammad SM, Turney PD (2013) Nrc emotion lexicon. 2013, NRC technical report
72. Hussein DMEDM (2018) A survey on sentiment analysis challenges. *J King Saud Univ Eng Sci* 30(4):330–338
73. Saad MK (2010) The impact of text preprocessing and term weighting on arabic text classification. *Comput Eng Islam Univ, Gaza*
74. Zerrouki T, Balla A (2017) Tashkeela: novel corpus of Arabic vocalized texts, data for auto-diacritization systems. *Data Brief* 11:147