

A Concept-based Sentiment Analysis Approach for Arabic

Ahmed Nasser¹ and Hayri Sever²

¹Control and Systems Engineering Department, University of Technology, Iraq

²Department of Computer Engineering, Çankaya University, Etimesgut

Abstract: *Concept-Based Sentiment Analysis (CBSA) methods are considered to be more advanced and more accurate when it compared to ordinary Sentiment Analysis methods, because it has the ability of detecting the emotions that conveyed by multi-word expressions in language. This paper presented a CBSA system for Arabic language which utilizes both of machine learning approaches and concept-based sentiment lexicon. For extracting concepts from Arabic, a rule-based concept extraction algorithm called semantic parser is proposed. Different types of feature extraction and representation techniques are experimented among the building processes of the sentiment analysis model for the presented Arabic CBSA system. A comprehensive and comparative experiments using different types of classification methods and classifier fusion models, together with different combinations of our proposed feature sets, are used to evaluate and test the presented CBSA system. The experiment results showed that the best performance for the sentiment analysis model is achieved by combined Support Vector Machine-Logistic Regression (SVM-LR) model where it obtained a F-score value of 93.23% using the Concept-Based-Features+Lexicon-Based-Features+Word2vec-Features (CBF+LEX+W2V) features combinations.*

Keywords: *Arabic Sentiment Analysis, Concept-based Sentiment Analysis, Machine Learning and Ensemble Learning.*

Received December 13, 2017; accepted July 29, 2019

<https://doi.org/10.34028/iajit/17/5/11>

1. Introduction

Currently, Sentiment Analysis (SA) or opinion mining is considered as one of the most rapidly emerging research areas due to the immediate need of processing the opinionated web contents coming from social networks and web blogs. SA is the task of determining the sentiment polarity of textual contents i.e., SA determines whether the emotions that expressed by a specific piece of text, is positive, negative or neutral [22].

Concept-Based (CB) SA methods are superior to standard word-level SA methods because they consider the meanings of multiple word expressions. CB SA approaches are concentrating on the semantic analysis of the textual contents through using semantic networks such as (SenticNet) and web ontologies, in order to extract the concepts that associated with the natural language opinions [3, 11].

CB emotion analysis is taking steps away from methods that use blind keyword and word co-occurrence frequencies, based on ontologies or semantic networks. The CB emotional analysis provides a better understanding of texts and offers a significant enhancement in the performance of the model. CB approaches can also detect complex emotions [9, 23].

The first step to Concept analysis was made by Wille [36] when he presented a "Formal Concept Analysis" (FCA) which is a mathematical model used for analysing and visualization data (configuration,

analysis, and visualization) and it is based on the concept of duality known as Galois connection [30]. Formal concepts are considered as formal summaries which involve clusters of data assets and their properties. Conceptual patterns are the type of conceptual structures which are consist of objects with their attributes that belong to specific areas. They are formed by specifying the objects and then their relations are demonstrated. The Fuzzy Formal Concept Analysis (FFCA) approach presented in [24] showed a great success in addressing the uncertainty information issues.

In [24] an FFCA based classification framework is proposed to classify document based on its conceptual summaries. The classification model is trained based on concepts using FFCA method. Thus, they intended to reduce the uncertainties that are affects the classifier performance. They have studied the polarity datasets of benchmark test bed (Reuters 21578) and two views on film and eBook interpretations. They have achieved good results in all data sets and have proved that the noisy drop sensitivity ability is good.

The work presented by Kontopoulos *et al.* [21] have adopted the FCA approaches for constructing an ontology field model. They used an ontology-based technique from their Twitter posts to make a more effective SA (by dividing each tweet into view sets tailored to the topic. They have worked on Smartphone spaces. The architectural views they use give a more detailed analysis of their posts. This also makes it

possible to distinguish the specific characteristics of the subject from the scores given to the subjects.

One of recently developed CB SA approaches is called pSenti and presented in Mudinas *et al.* [28]. This system is integrated learning-based approaches with data dictionary-based Opinion Mining (OM). The authors claim that the pSenti system has acquired a high emotion polarity classification performance in term of accuracy. At the same time, pure data dictionary has been compared with base systems in order to find emotion strength. They have tested the pSenti system using IMDB movie reviews and CNET software reviews datasets and they showed that pSenti has performed better than most current system-like hybrid approaches such as sentistrength.

Cambria *et al.* [10] have introduced senticnet. They have developed senticnet which act as a semantical link between concept-level emotion and natural word-level language data. They have built their systems with Sentic computation which is an integrated framework that taking the advantage of semanticweb and Artificial Intelligence (AI). Nowadays the social media and internet become a very simple and effective platform for the people for expressing their emotions and opinions through written text. The need of capturing the opinion of the public has raised due to the exceptional range of benefits that, include marking, business management, and financial forestation. However, mining opinion from languages is a very complex task because of its need a deep and complete understanding of the rules of the language. Conventional SA approaches are mainly dependent on the parts of the text in which opinions are expressed, based on features such as words co-occurrence frequency, keywords, and terms polarity. However, because these syntactical approaches are not relying on the natural language semantic and effective information of the text, these approaches are not efficient in detecting complex emotions.

CB approaches [3] are relying on the semantic and effective information that associated with the natural language opinions, which are represented as the concepts. CB SA approaches utilize the semantic networks and web ontologies for analysing the textual contents semantically.

This CB SA method is considered to be superior to the other ordinary sentiment analysis method because it's able to detect the emotions that conveyed by multi-word expressions concepts [3, 9]. Rather than gathering separated opinions, concepts-based analysis enables a comparative fine grind feature-based analysis. Common and common-sense can be considered as the key that enables feature spotting and polarity detection and it also necessary for dismantling the language into sentiment. Approaches of concept based sentiment analysis emphasize the effective knowledge-based resources such as wordnet [26], sentiwordnet [13] and senticnet [9].

2. Related Works in Sentiment Analysis for Arabic Language

There are many supervised and unsupervised approaches in the literature deal with the SA of the arabic language which are used to achieve the SA task in document-level or sentence-level [19]. In the supervised approach or the corpus-based approaches involves the generating of a sentiment decision model based on using an annotated sentiment corpus for training a different types of Machine Learning (ML) classification approaches such as K-Nearest Neighbours (KNN), Naive Bayes (NB), Decision Tree (D-Tree), Support Vector Machine (SVM), and etc., The alternative unsupervised approach or Lexicon Based (LB) approaches use a sentiment specific dictionaries in order to identify the polarity of a text based on the sentiment polarity of the individual words used in that text [2].

Rushdi-Saleh *et al.* [31] proposed a document-level supervised SA approach. They generated an arabic opinion corpus called OCA using the online movies reviews. For identifying the sentiment polarity, they used two types of ML classification methods which are NB and SVM. To extract the features from the Arabic documents they used various feature extraction methods based on n-gram representations and two different feature weighting techniques based on "Term Frequency" (TF) and "Term Frequency-Inverse Document Frequency" (TF-IDF).

Shoukry and Rafea [34] used sentence-level supervised SA approach for the arabic language by collecting the required data for SA from twitter. They applied two different feature extraction methods based on using bigrams and unigrams and TF weights together with NB and SVM ML-based classifier for building their proposed approach.

Mountassir *et al.* [27] three different solutions were proposed for solving the unbalancing issue in the datasets that used for SA. These methods include; "eliminate by clustering", "eliminate similar", and "eliminate farthest". In addition to that, they built a supervised approach for document-level arabic SA based on different types of ML classification methods such as KNN, NB, and SVM. They used a binary weighting which is based on term presence where the documents are considered as bags-of-words. Two types of imbalanced of Arabic and English corpus were used for evaluating their system, the first one consists from Arabic movie reviews that collected from "Al-Jazeera's website" and the second one consists from English product reviews and collected from the SINAI.

Duwairi *et al.* [18] used a supervised SA approach for tweets in the Arabic language. The authors generate a large dataset form tweeter and Facebook comments in different domains and manually tagged the polarity for each tweet and comment in the dataset. They used

three different ML-based classifiers such as NB, KNN, and SVM, as sentiment classification method.

Duwairi et al. [17] are also proposed a supervised learning approach for SA of tweets written using Arabizi (writing arabic using latin letters). They used rule-based method for converting Arabizi tweets to Arabic. Then the using crowdsourcing for assigning the sentiment polarity to each tweet to generate the dataset which used to build SA framework using two different classification techniques such as NB and SVM.

3. Structure and Organization of SenticNet

Senticnet is a CB sentiment lexicon that can be considered as one of the important resources that can be used for building a concept-based SA system. Senticnet use graph mining and multidimensional scaling to reduce the gap between the word and the opinions that covered by the word in natural language. Many applications have been developed by employing senticNet. These applications can be exploits in many fields such as an analysis a considerable amount of social data, human and computer interactions [10]. Senticnet-v3 consist of a 30k single and multi-word concepts while senticnet-v4 contains 50k of concepts. SenticNet provides different information about each concept, this information includes [12];

- Polarity which is a float number in the range between-1 to 1 that represents the sentiment score of the input concept.
- Five different single or multi-word senses semantically related to the input concept.
- Four different values that represent the diminutions of the hourglass emotion for the input concept.

4. Constructing Arabic Sentiment Lexicon (Ar-SenticNet)

The task of constructing the Arabic version of senticnet CB sentiment lexicon is consisting of two stages as shown in Figure 1.

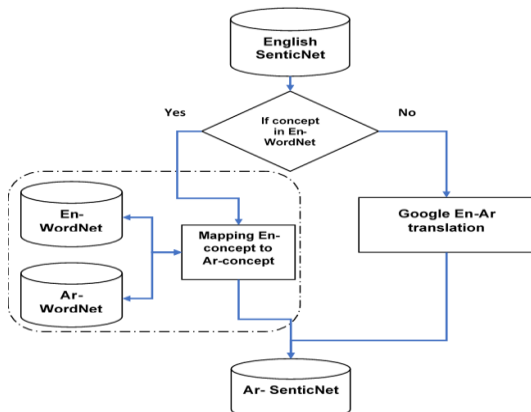


Figure 1. The proposed approach for translating English SenticNet to the Arabic language.

The first stage involves the process of translating each concept found in the English version of senticnet and the second stage is involving the extension process of the translated Arabic version of the senticnet, as following;

- The Translation Stage: the translation process of the senticnet conceptual lexicon to the Arabic language is done in two phases as shown in Figure 1. WordNet is considered as one of the resources that used to build the senticnet, so that in the first place we used a cross-language translation to translate SenticNet concepts to Arabic based on the mapping of both English wordnet [26] and Arabic wordnet [7]. The English concept that is required to be translated into Arabic is firstly searched in English WordNet and if it's found in the English WordNet, then a mapping between English and Arabic WordNet is used to obtain the Arabic translation of this concept. The second phase, in case, that the concept that required to be translated into Arabic is not found in the English wordnet, then the concept is translated into Arabic using Google Translation Application Programming Interface (API). Some examples of SenticNet concept translation to Arabic is shown below;

SenticNet--> trip_up ---->En-WordNet (found)--- mapping --- Ar-WordNet ---> أَمْسَكَ
 SenticNet--> care ----> En-WordNet (found)--- mapping --- Ar-WordNet ---> اِهْتَمَّ
 SenticNet--> catch_fire ----> En-WordNet (not found) ---- Google translate ----> اشتعل
 SenticNet--> long_trip ----> En-WordNet (not found) - --- Google translate ----> رحلة طويلة

- The Extension Stage: each concept with SenticNet has a set of senses called semantics. The aim of the extension process is to extend this set of senses by adding more senses which can be obtained from WordNet. The translated version of SenticNet is extended as following;

1. Searching all concepts that are translated into the Arabic in Ar-WordNet.
2. Obtaining the synonym sets for each concept that found in Ar-WordNet.
3. Adding these synonyms sets to Ar-SenticNet to extend the senses set for the concepts.

An example of the concept extension by adding sense from Ar-WordNet, is shown below.

“قائمة الحساب”----->Ar-WordNet----->(found)----->get synset----> [الحساب, حساب, كشف, تقدير, عد, عداد, حسبة, تعداد, احتساب, فاتورة]

We applied this translation and extension approach on the recently released senticnet-v4 which is consist of 50k English concepts to generate the Arabic version of senticnet as following;

1. 9.4k of senticnet-v4 English concepts were translated into Arabic using wordnet mapping.
2. Another 30k of English concepts were translated by using Google translate method. This resulted in a 39.4k of Arabic concepts.
3. Then by applying the extension process, 11.2k of the translated concepts were found in Arabic WordNet and the obtained synonyms set for these concepts are added another 9k sense to extend the number of concepts in Arabic SenticNet.

This resulted in a total of 48k of Arabic concepts and thus the Ar-SenticNet is generated.

5. Our Approach for CB Arabic Sentence Level Sentiment Analysis

The proposed architecture of our ML approach for Sentence-level CB Arabic SA system is shown in Figure 2.

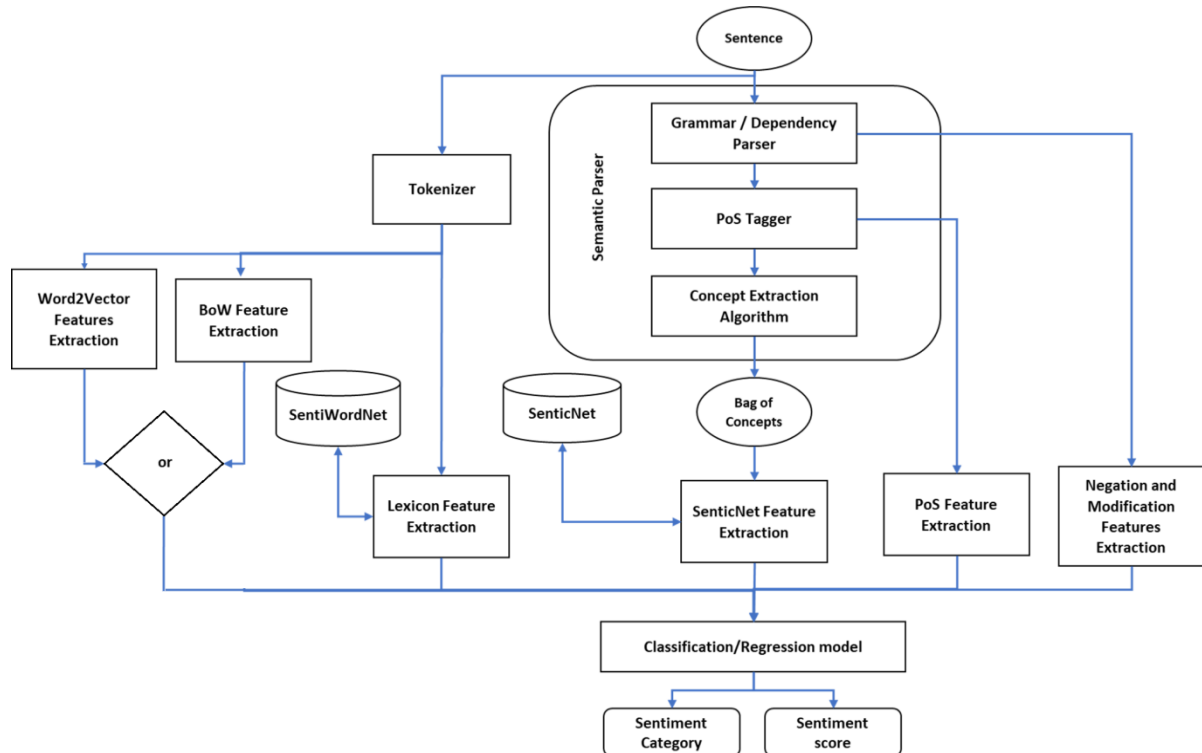


Figure 2. The architecture of the proposed ML approach for Sentence-level CB Arabic SA system.

The architecture of our CB Arabic Sentence-level SA approach shown in Figure 2 involves different stages such as (concept extraction, feature extraction and the ML algorithms for the SA task). These stages are explained as follows;

5.1. Semantic Parser (SP)

The concepts can be simply defined as the single or multi-word expression that carries the meaning of a phrase or sentence. Semantic parsing is considered as the task of extraction the concepts for the sentence based on its grammatical structure. Concept extraction process involves fragmenting and partitioning the sentence into a noun and verb clauses then form a candidate list of words that match the grammatical rules of the concept in those parts.

- Step 1 “Stanford Arabic parser”¹ is used to extract noun and verb phrases from the sentence.
- Step 2 “Stanford Arabic Tagger”¹ tool is used to assign the part of speech tags to each word found in

the noun and verb phrases. Figure 3 shows an example of the parse tree of an Arabic sentence after applying the steps one and two.

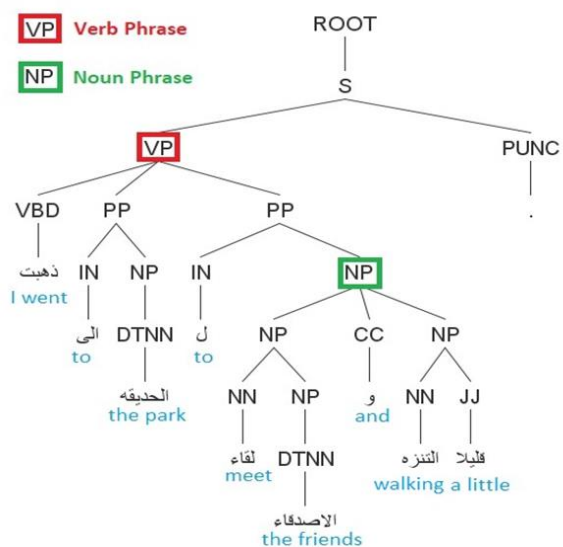


Figure 3. An example of Arabic sentence parse tree with PoS tags which generated by using both of Stanford Arabic Parser and Tagger tools.

¹<https://nlp.stanford.edu/projects/arabic.shtml>

- For the noun clauses, the algorithm looks at the word sequence of each (Bigram) word pair and adds the words matching the following rule (pattern) to the candidate list of concepts.
- [Noun+Noun] → add [Noun+Noun] pattern to the list of candidate concepts.
- [Noun+Adjective] → add [Noun+Adjective] pattern to the list of candidate concepts.
- [Adjective+Noun] → add [Adjective+Noun] pattern to the list of candidate concepts.
- [Noun+Stopword] → add only the [Noun] to the list of candidate concepts.
- [Stopword+Adjective] → add only [Adjective] to the list of candidate concepts.
- For each one of the verb clauses the algorithm adds words that match the [Verb+Object] rule to the candidate list of concepts. In the first step, in order to determine the Object that related to the Verb in the sentence, the Stanford dependency parser “dependency analyser”¹ is used. Stanford dependency parser is used to specify and identify grammatical dependency relationships among the words in the phrase. The Arabic language dependency dataset² is used to train the Stanford dependency parser in order to use it for the Arabic dependency analysis.

After analysing the sentence with the “Stanford dependency parser”, the words that match the [Verb+Object] rules are added to the candidate list of concepts.

In some cases, the “Stanford dependency parser” may not be able to identify the object because of the limitations in the Arabic dependency dataset. If this is the case, the grammar structure of the standard Arabic sentence ([Verb + Subject] + Object) is used to find the object in the sentence [1]. When the Object is identified then the Object added to the candidate concepts list with the Verb related to it. The generated final list of the candidate concepts is called Bag of Concepts.

Our SP concept extraction algorithm is tested against randomly selected 100 Arabic sentences with manually extracted concepts. For these 100 Arabic sentences, the concepts that extracted by the proposed concept extraction algorithm are compared with the manually extracted concepts. The comparison results show that the concept extraction algorithm is achieved an accuracy value of 97% for concept extraction compared to the manual method. Table 1 shows an example of Arabic sentences and the concepts extracted from these sentences using our concept extraction algorithm.

Table 1. An example of concepts that extracted from Arabic of sentences using our SP concept extraction algorithm.

1	Arabic sentence	إن أسعار الذهب بالغة الحساسية لتحركات أسعار الفائدة العالمية
	English translation	Gold prices are very sensitive to movements of the global interest rate.
	Grammatical parse	(ROOT (S (VP (VBP (إن) (NP (NN أسعار) (NP (DTNN الذهب)))) (ADJP (JJ بالغة) (NP (DTNN الحساسية)))) (PP (IN ل) (NP (NNS تحركات) (NP (NN أسعار) (NP (DTNN الفائدة) (DTJJ العالمية)))))))))
	Extracted concepts	سعر ذهب الحساسية لتحركات تحركات أسعار الفائدة العالمية حساسية لحركة سعر الفائدة
2	Arabic sentence	يمكن أن يكون المحيط شياً في غاية التعقيد
	English translation	The ocean can be a very complicated thing.
	Grammatical parse	(ROOT (S (VP (VBP (يمكن) (NP (DTNN أن) (S (VP (VBP (يكون) (NP (DTNN المحيط) (DTJJ (أمرا) (PP (IN في) (NP (NN غاية) (NP (DTNN التعقيد)))))))))))))
	Extracted concepts	المحيط المحيط شيء شيء معقد شيء غاية التعقيد معقد المحيط

5.2. Feature Extraction and Representations

Since ML approaches are considered as the main core of our proposed Sentence-level CB Arabic SA approach, which are responsible for identifying and deciding the sentiment polarity of the input sentence so that the input sentence must be transformed into a set of numerical features that can be useful for the ML algorithm. In this work, we present and exploit a different feature extraction and representation techniques, to extract a variety of features sets from the input sentence and then fed them as input to ML decision model. These feature set are CB features, lexicon-based features, Bag of Word features and Word2Vector features.

1. The CB Features (CBF) Includes;

- SenticNet features (The number of concepts extracted from the sentence and found in our generated Arabic SenticNet, the summation of the extracted concepts scores which are obtained from Arabic SenticNet).
- Part of Speech (PoS) features (The number of nouns, adjectives, and adverbs found in the sentence).
- Modification features (This binary feature is set to 1 if the sentence has any word modified by an adverb, adjective, or noun otherwise it is set to 0).
- Negation features (The negation binary feature determines whether there is any negation in the sentence).

2. The Lexicon Based Features (LEX) Includes;

- Lexicon features (Positive words number, Negative words number, Positive words number divided by the negative word number, the sum of

²<http://universaldependencies.org/>

the positive scores and the sum of the negative scores). The version of Arabic SentiWordNet that called ArSneL [4] is used to extract these features from the sentence.

3. The Bag-of-Word Features (BoW) Includes;

- Bag-of-Word features (the sentence is represented using either Uni-grams or Bigrams features as a vector, and these features are weighed using TF-IDF method).

4. The Word2Vector Features (W2V) Includes;

- Word2vector features (each word within the sentence is transformed into a real-valued 300-dimensional vector, the Word2Vector features are generated by the aggregation of the vectors of each word in the sentence) [5]. Word Vectors can be obtained from the learned Word2Vec model presented by Mikolov *et al.* [25]. We used our large-scale corpus GLASC to train and generate the Word2Vec model.

5.3. Corpus Description

Our GLASC corpus refers to “GDELT Large-scale Arabic Sentiment Corpus” which presented in [29] and it is built using online Arabic news articles and metadata provided by the Bigdata resource GDELT (Global Database of Events, Language, and Tone) [8, 23]. The process of generating GLASC corpus is illustrated in Figure 4. Our corpus consists of a total of 620,082 Arabic news articles divided into three categories (225,397 Positive, 266,376 Negative and 218,309 Neutral articles).

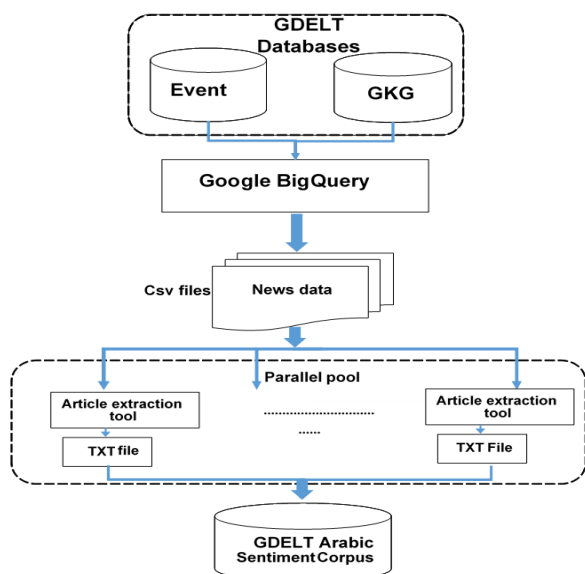


Figure 4. Our GLASC corpus generation process.

5.4. ML Algorithms for SA

For the CB SA, we utilized ML algorithms where the SA task of a sentence can be considered as classification problem. In this work, ML classification

approaches are considered to identify the sentiment category of the input sentence.

For the sentiment classification task, we employ four different ML classifiers such as (SVM [20], Hidden Markov Model (HMM) [35], NB [15], and LR [16]) to generates the sentiment classification model. Moreover, different versions of combined classifiers methods such as (Support Vector Machine-Logistic Regression (SVM-LR), Support Vector Machine Naïve Bayes (SVM-NB), and Support Vector Machine-Hidden Markov Model (SVM-HMM)) which are ensembled using stacking technique [32], are also taken to our considerations. These classification algorithms can be trained using a sentence-based dataset that is can be Generated From Our Arabic Large-Scale SA Corpus (GLASC).

6. CB Approach for Arabic Sentence-level Sentiment Analysis Evaluation

In this section, we performed comprehensive comparative experiments for evaluating our proposed CB Sentence-level Arabic SA system which is based on ML classification and regression approaches. Our goal with this evaluation experiments is to determine and specify the best performing ML classification models to use them for the proposed SA system.

6.1. Measure the Coverage of the Translated Arabic SenticNet

In order to evaluate the quality of our translated Arabic SenticNet CB sentiment lexicon, we calculate its coverage over our GLASC corpus. The coverage “Cov” of Ar-SenticNet can be calculated as following;

$$Cov = \frac{|[Arabic_senticNet] \cap GLASC|}{|GLASC|} \quad (1)$$

Using the formula described above we calculate the coverage of our Ar-SenticNet which is generated using our proposed translation and extension process. The Ar-SenticNet contains a total number of concepts of 48,343. Based on this coverage calculation formula, the Ar-SenticNet has obtained a 73.3% coverage over our GLASC corpus.

The coverage results show that the Ar-SenticNet has obtained a sufficient coverage which can demonstrates its effectiveness by covering a wide range of Arabic concepts.

6.2. Dataset and Feature Exaction

To build the ML classification and regression models for our proposed CB SA system, a dataset is generated from our GLASC corpus. This dataset consists of a total of 1750 sentence organized as (594 negative, 585 positive and 571 neutral) and each sentence contains an average number of 30 words. We applied the feature

extraction methods explained in Section 3.2, on this dataset to generate different feature vectors which are then used to train and evaluate the system. These distinctive features such as CB features CBF, Lexicon based features, Bag of Word features and Word2Vector features are extracted from the dataset. We intended to use different feature combinations in order to evaluate our system. These combinations of features are shown and described in Table 2 below.

Table 2. The proposed feature combinations and its descriptions.

Features	Description
CBF	Using only CB Features
CBF+LEX	Using CB Features together with the Lexicon Based Features
CBF+W2V	Using CB Features together with the Word Vector Features
CBF+ BoW_Uni	Using CB Features together with the Bag of Words Features that generated using Unigrams and TF-IDF weighting
CBF+ BoW_Bi	Using CB Features together with the Bag of Words Features that generated using Bigrams and TF-IDF weighting
CBF+LEX+BoW_Bi	Using CB Features combined with the Lexicon Based Features and the Bag of Words Features that generated using Bigrams and TF-IDF weighting.
CBF+LEX+W2V	Using CB Features combined with the Lexicon Based Features and the Word Vector Features

6.3. ML Models Generation and Evaluation

In order to evaluate the performance of our proposed ML-based sentiment classification and regression models, we considered using 10-fold cross-validation method to calculate the classification accuracy and F score for the ML sentiment classification models. These evaluation measures are explained as following, [6, 14];

$$Accuracy = \frac{1}{c} \sum_{i=1}^c \frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i} \quad (2)$$

$$F - score = \frac{1}{c} \sum_{i=1}^c \frac{2TP_i}{2TP_i + FP_i + FN_i} \quad (3)$$

Where c refers to the number of classes and True Positive (TP) and True Negative (TN) refers to the number of instances that correctly classified by the model, False Positive (FP) and False Negative (FN) refers to the number of instances that miss-classified by the model.

6.3.1. Experiments with Base Learners

In order to generate our classification model, we examined four different ML classifier algorithms such as SVM, LR, HMM, and NB. We performed a 10-fold cross validation separately on these four classifiers using different sets of feature combinations and calculated the accuracy and F-score classification model performance evaluation metrics. The obtained accuracy and F-score results for each classifier using distinctive features combinations are reported in Table 3.

Table 3. The classification Accuracy and F-score for base learners using different features combinations (Acc is accuracy and F-sc is F-score).

			Classifiers			
			SVM	LR	HMM	NB
Features	CBF	Acc	0.7245	0.7188	0.7375	0.6734
		F-sc	0.7099	0.7058	0.7284	0.6561
	CBF+LE X	Acc	0.7735	0.7551	0.7478	0.7088
		F-sc	0.7722	0.7274	0.7102	0.6996
	CBF+W2 V	Acc	0.8979	0.8607	0.8852	0.8206
		F-sc	0.8924	0.8533	0.8849	0.8200
	CBF+Bo W_Uni	Acc	0.8525	0.8457	0.8547	0.8013
		F-sc	0.8491	0.8352	0.8527	0.7787
	CBF+Bo W_Bi	Acc	0.8827	0.8880	0.8785	0.8220
		F-sc	0.8771	0.8822	0.8738	0.8210
	CBF+LE X+BoW_ Bi	Acc	0.8930	0.8921	0.8855	0.8562
		F-sc	0.8929	0.8895	0.8853	0.8462
	CBF+LE X+W2V	Acc	0.9104	0.9035	0.9043	0.8674
		F-sc	0.9089	0.9000	0.9015	0.8670

From the results in Table 3 it can be shown that the CBF+LEX+W2V features combination provides the best classification performance of all the classification models that used. SVM classifier provided its best performance of 90.89% F-score using CBF+LEX+W2V features combinations over all other classifiers and other features combinations sets. CBF+LEX+BoW-Bi features combination also provides a very good classification performance for all classifier model less than about only 1.5% form the maximum performance that provided by CBF+LEX+W2V features combination. By using only CBF the best result is obtained by HMM classifier with 72.84% of F-score. Combining CBF with other features such as LEX, BoW, and W2V increases the classification performance for all classifier models used.

6.3.2. Experiments with Classifier Model Fusion

In order to enhance and improve the classification performance of our classification model we also considered using classifier model fusion approach. We used a different combination of classifiers methods such as (SVM+LR, SVM+NB, and SVM+HMM) as level-0 models. For the level-1 model, we used a multilayer perceptron MLP as meta-classifier to combine the decisions of the level-0 classifier models. Each one of these fused classifier models were individually evaluated using 10-fold cross-validation based on distinctive features sets and the obtained values of classification accuracy and F-score is reported in Table 4 below.

Table 4. The classification Accuracy and F-score for combined learner using stacking method by using distinctive features combinations (Acc is accuracy and F-sc is F-score).

		Classifiers			
		SVM-LR	SVM-NB	SVM-HMM	
Features	CBF	Acc	0.7356	0.7264	0.7482
		F-sc	0.7327	0.7264	0.7456
	CBF+LEX	Acc	0.7784	0.7755	0.7871
		F-sc	0.7783	0.7682	0.7862
	CBF+W2V	Acc	0.9042	0.9134	0.8993
		F-sc	0.9022	0.9126	0.8991
	CBF+BoW_Uni	Acc	0.8651	0.8739	0.8592
		F-sc	0.8582	0.8737	0.8591
	CBF+BoW_Bi	Acc	0.8911	0.8878	0.8935
		F-sc	0.8859	0.8872	0.8914
	CBF+LEX+BoW_Bi	Acc	0.9005	0.8965	0.9079
		F-sc	0.8968	0.8959	0.9059
CBF+LEX+W2V	Acc	<u>0.9396</u>	0.9164	0.9340	
	F-sc	<u>0.9392</u>	0.9131	0.9323	

When comparing the results of combined classifier models with the result of base learners, it is shown that the combined models have an impact on increasing the classification performance. Similar to based learner results the best classification performance is obtained using CBF+LEX+W2V features combinations. For all used features combination, the best performance is achieved by SVM-LR combined classification. SVM-LR model proved its best classification performance of 93.92% F-score using CBF+LEX+W2V features combinations. The best performance using only CBF is achieved by SVM-HMM model with F-score of 74.56%. Using CBF combined with the other features such as LEX, BoW and W2V increases the classification performance for all classifier models used. The result in Figure shows that the SVM-HMM fused classification model has the best performance over the other combined classifiers where it's achieved an average value of F-score of 85.99% for all features combination.

Since we used different features combinations in order to evaluate our sentiment classification system, we investigate the impact of using different feature combination sets on the classification performance. In order to calculate the percentage of performance improvement using these different features combinations, since CBF features are the common point in all our features combinations, we first selected the CBF as baseline for our experiment, then we calculate the percentage of improvement in F-score for each classifier/classifier-combinations, when a different feature is combined with baseline features. The results in Figure 5, shows the impact of combining different features with CBF features on improving the classification accuracy for each classifier.

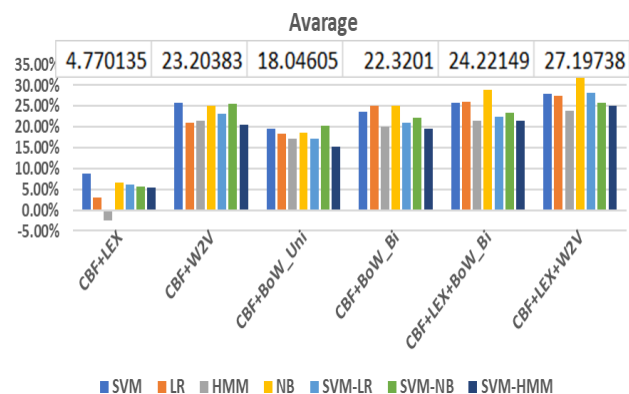


Figure 5. The percentage of improvement in F-score for each classifier when using the other features are combined with CBF baseline features.

Based on the result shown in Figure 5 combining LEX and W2V features with the CBF baseline features, improve the average accuracy of the classifiers by 27.2% compared to accuracy obtained by using only CBF baseline features. This mean combining concept-based SA approach with the ordinary SA approaches can improve the accuracy of the SA systems.

We can summarize the experiments result, as follows: Using the baseline CBF features individually results in a poor classification performance for all used classification models. Combining various features such as (LEX, BoW-Uni, Bow-Bi, W2V) with the baseline CBF features results in improving the classification performance for all used classification models. The maximum improvement in classification performance for all used classifiers features is achieved using (CBF+LEX+W2V) features combinations. Combining LEX features with the CBF+W2V and CBF+BoW features combinations resulted in an improvement in classification performance for all used classifiers. However, combining only LEX features with the CBF features resulted in a less improvement in classification performance compared to the improvement achieved by combining only W2V or BoW features with the CBF features. Combining BoW based features with the baseline CBF features is also achieved a responsible improvement in classification performance. The maximum classification performance is achieved by SVM base learner classifier for the all different features combinations followed by HMM then LR and NN classifier. Using classifier model fusion by stacking method is able to improve the accuracy of the all combined (fused) classification models rather than using each single classifier model separately. The maximum classification performance is achieved by using SVM+LR classifier fusion model with the highest F-score value of 93.92%, so that this method is considered for the generation the sentiment classification model that used in our proposed concept-based Sentence-level Arabic SA system.

7. Conclusions

In this work we presented a concept-based sentiment lexicon for the Arabic language which is generated by translating the English version of the concept-based sentiment lexicon SenticNet to Arabic using two-way translation and extension process. The translation process is utilizing the English-Arabic cross-language mapping that provided by WordNet and the Google translation service. Then the translated Arabic concept-based sentiment lexicon is extended by adding more senses which obtained from Arabic WordNet.

We applied this translation and extension approach on the recently released SenticNet-v4 which consist of 50k of English concepts and resulted in the Ar-SenticNet with 48k Arabic concepts.

In order to evaluate the quality of our translated Arabic SenticNet concept-based sentiment lexicon, we calculate it coverage over our GLASC corpus based on coverage calculation formula that described in Section. Based on the calculation that we did, the Ar-SenticNet is obtained a 73.3% coverage over our GLASC corpus and this means that the Ar-SenticNet provide very good cover to most of the concepts that found in our large-scale GLASC corpus.

We also build a concept-based sentiment analysis system for Arabic Sentence-level sentiment analysis using our previously mentioned Ar-SenticNet concept-based sentiment lexicon and a variety ML approaches. For extracting the concept from the Arabic sentence, we proposed and performed a rule-based concept extraction algorithm called semantic parser. In order to generate the candidate concept list for an Arabic sentence, this semantic parser utilizes a variety of freely available grammatical and morphological analysis tools for the Arabic language beside the grammatical rules of the Arabic concepts.

We also presented and used different types of feature extraction and representation techniques for building the concept-based Sentence-level Arabic sentiment analysis system. These techniques used to extract various feature sets from the input sentence, which used to build the ML decision model. These feature set are concept-based features CBF, lexicon-based features LEX, Bag of Word features BoW and Word2Vector features W2V.

For building the ML-based decision model the concept-based Sentence-level Arabic sentiment analysis system we used different types of ML classification method such as (SVM, HMM, NB and LR). In order to improve the classification performance, we also used classifier fusion method for combining classification models such as (SVM-HMM, SVM-NB, and SVM-LR). For training these ML-based models we generated a sentence-based dataset form our GLASC corpus and carried out a comprehensive and comparative experiments using a different combination of the feature sets that mentioned earlier

with the baseline concept-based features CBF. The features combinations that we used are (CBF, CBF+LEX, CBF+W2V, CBF+BoW_Uni, CBF+BoW_Bi, CBF+LEX+BoW_Bi and CBF+LEX+W2V).

Our experiment results show that the best performance for the classification model is achieved by using SVM classifier which obtained an F-score value of 90.89% using CBF+LEX+W2V features combinations where the combined SVM-LR model is obtained better classification performance of 93.23% F-score using the same CBF+LEX+W2V features combinations.

For the future works; we are considering using an approach similar to the one used in [33] for expanding both of the Arabic sentiment lexicon SentiWordNet (ArSenL) and the Arabic concept-based sentiment lexicon SenticNet, using our large-scale corpus GLASC. We are also considering using rule-based SA approaches together with concept-based SA approaches, which can lead to increase in the precision and accuracy of SA by using a language dependent grammatical rule.

References

- [1] Abdelali A., Cowie J., and Soliman H. S., "Arabic Information Retrieval Perspectives," in *Proceedings of 11th Conference on Natural Language Processing, Journes d'Etude sur laParole-Traitement Automatique des Langues Naturelles*, Fez, pp. 391-400, 2004.
- [2] Abdulla N., Ahmed N., Shehab M., and Al-Ayyoub M., "Arabic Sentiment Analysis: Corpus-Based and Lexicon-Based," in *Proceedings of IEEE Conference on Applied Electrical Engineering and Computing Technologies*, Amman, pp. 1-6, 2013.
- [3] Agarwal B., Poria S., Mittal N., Gelbukh A., and Hussain A., "Concept-Level Sentiment Analysis with Dependency-Based Semantic Parsing: A Novel Approach," *Cognitive Computation*, vol. 7, no. 4, pp. 487-499, 2015.
- [4] Badaro G., Baly R., Hajj H., Habash N., and El-Hajj W., "A Large Scale Arabic Sentiment Lexicon for Arabic Opinion Mining," in *Proceedings of EMNLP on Arabic Natural Language Processing*, Doha, pp. 165-173, 2014.
- [5] Bilgin M. and Köktaş H., "Sentiment Analysis with Term Weighting and Word Vectors," *The International Arab Journal of Information Technology*, vol. 16, no. 5, pp. 953-959, 2019.
- [6] Bishop M., *Pattern Recognition and Machine Learning*, Springer, 2006.
- [7] Black W., Elkateb S., Rodriguez H., Alkhalifa M., Vossen P., Pease A., and Fellbaum C., "Introducing the Arabic Wordnet Project," in *Proceedings of 3rd International Wordnet*

- Conference, South Jeju Island, pp. 295-300, 2006.
- [8] Bodas-Sagi D. and Labeaga J., "Using GDELT Data to Evaluate the Confidence on the Spanish Government Energy Policy," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 3, no. 6, pp. 38-43, 2016.
- [9] Cambria E. and White B., "Jumping NLP Curves: A Review of Natural Language Processing Research," *IEEE Computational Intelligence Magazine*, vol. 9, no. 2, pp. 48-57, 2014.
- [10] Cambria E., Havasi C., and Hussain A., "SenticNet 2: A Semantic and Affective Resource for Opinion Mining and Sentiment Analysis," in *Proceedings of the 25th International Florida Artificial Intelligence Research Society Conference*, Marco Island, pp. 202-207, 2012.
- [11] Cambria E., Schuller B., Liu B., Wang H., and Havasi C., "Knowledge-Based Approaches to Concept-Level Sentiment Analysis," *IEEE Intelligent Systems*, vol. 28, no. 2, pp. 12-14, 2013.
- [12] Cambria E., Soujanya P., Rajiv B., and Björn W., "Senticnet 4: A Semantic Resource for Sentiment Analysis Based on Conceptual Primitives," in *Proceedings of 26th International Conference on Computational Linguistics*, Osaka, pp. 2666-2677, 2016.
- [13] Cambria E., Speer R., Havasi C., and Hussain A., "SenticNet: A Publicly Available Semantic Resource for Opinion Mining," in *Proceedings of AAAI Fall Symposium Series*, Washington, pp. 14-18, 2010.
- [14] Davis J. and Goadrich M., "The Relationship between Precision-Recall and ROC Curves," in *Proceedings of 23rd international conference on Machine learning*, Corvallis, pp. 233-240, 2006.
- [15] Dhande L. and Patnaik G., "Analyzing Sentiment of Movie Review Data Using Naive Bayes Neural Classifier," *International Journal of Emerging Trends and Technology in Computer Science*, vol. 3, no. 4, pp. 313-320, 2014.
- [16] Dreiseitl S. and Ohno-Machado L., "Logistic Regression and Artificial Neural Network Classification Models: A Methodology Review," *Journal of Biomedical Informatics*, vol. 35, no. 5, pp. 352-359, 2002.
- [17] Duwairi R., Faqeeh M., Wardat M., and Alrabadi A., "Sentiment Analysis for Arabizi Text," in *Proceedings of 7th International Conference on Information and Communication Systems*, Irbid, pp. 127-132, 2016.
- [18] Duwairi M., Marji R., Sha'ban N., and Rushaidat S., "Sentiment Analysis in Arabic Tweets," in *Proceedings of 5th International Conference on Information and Communication Systems*, Irbid, pp. 1-6, 2014.
- [19] Farra N., Challita E., Assi R., and Hajj H., "Sentence-Level and Document-Level Sentiment Mining for Arabic Texts," in *Proceedings of IEEE International Conference on Data Mining Workshops*, Sydney, pp. 1114-1119, 2010.
- [20] Joachims T., "Text Categorization with Support Vector Machines: Learning With Many Relevant Features," in *Proceedings of European Conference on Machine Learning*, Chemnitz, pp. 137-142, 1998.
- [21] Kontopoulos E., Berberidis C., Dergiades T., and Bassiliades N., "Ontology-Based Sentiment Analysis of Twitter Posts," *Expert Systems with Applications*, vol. 40, no. 10, pp. 4065-4074, Aug. 2013.
- [22] Korayem M., Crandall D., and Abdul-Mageed M., "Subjectivity and Sentiment Analysis of Arabic: A Survey," in *Proceedings of International Conference on Advanced Machine Learning Technologies and Applications*, Cairo, pp. 128-139, 2012.
- [23] Leetaru K. and Schrodt P., "GDELT: Global Data on Events, Location, and Tone," in *Proceedings of ISA Annual Convention*, San Francisco, pp. 1-49, 2013.
- [24] Li T. and Tsai C., "A Fuzzy Conceptualization Model for Text Mining With Application in Opinion Polarity Classification," *Knowledge-Based Systems*, vol. 39, no. Supplement C, pp. 23-33, and 2013.
- [25] Mikolov T., Sutskever I., Chen K., Corrado G. S., and Dean J., "Distributed Representations of Words and Phrases and their Compositionality," in *Proceedings of the 26th International Conference on Neural Information Processing Systems*, Lake Tahoe, pp. 3111-3119, 2013.
- [26] Miller A., Beckwith R., Fellbaum, C., Gross D., and Miller K J., "Introduction to Wordnet: An On-Line Lexical Database," *International journal of lexicography*, vol. 3, no. 4, pp. 235-244, 1990.
- [27] Mountassir A., Benbrahim H., and Berrada I., "Some Methods to Address The Problem of Unbalanced Sentiment Classification in an Arabic Context," in *Proceedings of Colloquium in Information Science and Technology*, Fez, pp. 43-48, 2012.
- [28] Mudinas A., Zhang D., and Levene M., "Combining Lexicon and Learning Based Approaches for Concept-level Sentiment Analysis," in *Proceedings of 1st International Workshop on Issues of Sentiment Discovery and Opinion Mining*, New York, pp. 1-8, 2012.
- [29] Nasser A. and Sever H., "A Large-Scale Arabic Sentiment Corpus Construction Using Online

- News Media,” *Journal of Engineering and Applied Sciences*, vol. 13, pp. 7329-7340, 2018.
- [30] Priss U., “Formal Concept Analysis in Information Science,” *Annual Review of Information Science and Technology*, vol. 40, no. 1, pp. 521-543, 2006.
- [31] Rushdi-Saleh M., Martín-Valdivia M., Ureña-López L., and Perea-Ortega J., “OCA: Opinion Corpus for Arabic,” *Journal of the Association for Information Science and Technology*, vol. 62, no. 10, pp. 2045-2054, 2011.
- [32] Ruta D. and Gabrys B., “An Overview Of Classifier Fusion Methods,” *Computing and Information systems*, vol. 7, no. 1, pp. 1-10, 2000.
- [33] Sağlam F., Sever H., and Genç B., “Developing Turkish Sentiment Lexicon for Sentiment Analysis Using Online News Media,” in *Proceedings of 13th International Conference of Computer Systems and Applications*, Agadir, pp. 1-5, 2016.
- [34] Shoukry A. and Rafea A., “Sentence-level Arabic Sentiment Analysis,” in *Proceedings of International Conference on Collaboration Technologies and Systems*, Denver, pp. 546-550, 2012.
- [35] Soni S. and Sharaff A., “Sentiment Analysis of Customer Reviews Based on Hidden Markov Model,” in *Proceedings of International Conference on Advanced Research in Computer Science Engineering and Technology*, New York, pp. 1-5, 2015.
- [36] Wille R., “Restructuring Lattice Theory: An Approach Based on Hierarchies of Concepts,” in *Proceedings of International Conference on Formal Concept Analysis*, Berlin, pp. 445-470, 1982.



Ahmed Nasser got his BSc degree from University of Technology Control and Systems Eng. Faculty, Baghdad Iraq in 2006, MSc degree in Computer Eng. from Istanbul University, Istanbul Turkey in 2012, PhD degree from Hacettepe University, Ankara Turkey Computer Eng., in 2018. His current interest on “Data Mining”, “Natural Language Processing” and “Machine Learning”.



Hayri Sever is currently working for Software Engineering Department at Cankaya University. He has received his BSc degree in computer science and engineering from Hacettepe University in Ankara, TR, MSc degree from Maine University in Orono, ME in 1991, PhD degree from Louisiana University in Lafayette LA, in 1995. His research areas are Knowledge Discovery in Databases, Multimedia Retrieval Models and Systems, Multimedia Systems, Uncertainty Reasoning, Business Process Management, Machine Learning, and Speech Analysis.