

# Semantic Sentiment Analysis of Arabic Texts

Sana Alowaidi, Mustafa Saleh, Osama Abulnaja  
Computer Science Department  
King Abdulaziz University  
Jeddah, Saudi Arabia

**Abstract**—Twitter considered as a rich resource to collect people's opinions in different domains and attracted researchers to develop an automatic *Sentiment Analysis* (SA) model for tweets. In this work, a semantic Arabic Twitter Sentiment Analysis (ATSA) model is developed based on supervised machine learning (ML) approaches and semantic analysis. Most of the existing Arabic SA approaches represent tweets based on the *bag-of-words* (BoW) model. The main limitation of this model is that it is semantically weak; where words considered as independent features and ignore the semantic associations between them. As a result, synonymous words that appear in two tweets are represented as different independent features. To overcome this limitation, this work proposes enriching the tweets representation with concepts utilizing Arabic WordNet (AWN) as an external knowledge base. In addition, different concepts representation approaches are developed and evaluated with naïve Bayes (NB) and support vector machine (SVM) ML classifiers on an Arabic Twitter dataset. The experimental results indicate that using concepts features improves the performance of the ATSA model compared with the basic BoW representation. The improvement reached 4.48% with the SVM classifier and 5.78% with the NB classifier.

**Keywords**—Arabic Sentiment Analysis; Twitter; Semantic Relations; Arabic WordNet; Machine Learning

## I. INTRODUCTION

Currently, Twitter is considered to be one of the most popular microblogs. It has allowed people to communicate, share comments, and express their opinions on almost all aspects of daily life at an increasing rate. Since analyzing huge volumes of opinionated text remains a formidable task, the high demand for automated sentiment analysis (SA) models became a necessity.

Sentiment analysis, which is also called opinion mining, is the computational study of people's opinions, sentiments, and attitudes about topics, entities, people, and events, that are expressed in texts [1]. It aims to assign a predefined sentiment class to online texts as negative, positive, or neutral. SA plays a substantial role in several domains such as financing, marketing, politics, and social.

One of the main approaches used to solve the SA problem is the supervised machine learning (ML) approach. In this approach, texts are represented by feature vectors which are used to train ML classifiers, such as naïve Bayes (NB) and support vector machines (SVMs), to infer a combination of particular features yielding a certain sentiment class. The resulting classifier model is then used to predict the sentiment class of the new un-annotated documents [1].

The performance of the SA model relied on the classifier algorithm and the text representation model. Various classifiers have been adopted for SA, but the challenging task is to engineer a set of powerful features to build a good representation model [1]. The vector space model (VSM) [2], also called the bag-of-words (BoW) model, is considered as a fundamental text representation model used in most ML approaches because of its simplicity and effectiveness. This model represents texts as a weighted features vector with words as basic features.

Many of the existing approaches in both the English and Arabic languages attempt to enhance the performance of the SA model by expanding the BoW model with different features such as word n-grams, POS tags, and stems. Also, new microblog features were proposed for Twitter data. However, the resulting representation models still suffer from a common limitation, they are semantically weak. The BoW model considers the words as independent features and ignores the semantic associations between the words. For example, it treats synonymous words as unrelated features. Moreover, in this model, only words that are explicitly mentioned in the training dataset are used to train the classifier, thereby ignoring the words in the testing documents that were not found in the training documents.

Recently, several studies, most of which were in the English language, have been proposed using a new semantic concepts representation model in various text mining (TM) fields including clustering [3], topic classification [4-6], and SA [7, 8]. Rather than representing the documents in their lexical space depending on BoW features, the semantic approach represents the documents in their semantic space as a set of concepts features extracted utilizing an external knowledge base (KB) such as WordNet (WN).

The Arabic language is a Semitic language which consists of 28 letters. It is a cursive language, in which word formation consists of connecting letters to each other. As opposed to the English language, Arabic writing starts from right to left, and has no capitalization.

The Arabic language is one of the fastest-growing languages on the web with about 168 million Arabic-speaking people using the Internet [9]. According to the Internet World Stat 2016 ranking [9], the Arabic language ranked in the top five languages used most on the Internet. However, while much SA research has been done for the English language, since it is a dominant language of science, little has been done for the Arabic language. The Arabic language poses a number of challenges, especially in regards to sentiment analysis. It not only that it has a very complex morphology compared to the

English language, but it is also a very derivational and inflectional language which makes morphological analysis a very complicated task [10, 11].

This paper presents an Arabic Twitter Sentiment Analysis (ATSA) model, a semantic sentiment analysis model for Arabic Twitter data using ML approaches. Unlike existing Arabic SA models which represent tweets texts in their lexical space based on BoW features, semantic concepts representation approach was proposed which aims to represent tweets in their semantic space by taking into account the semantic relationships between the words by utilizing the Arabic WordNet (AWN).

The rest of this paper is structured as follows. Section II examines previous related work. Section III describes the ATSA model. Section IV presents the proposed concepts representation approach in detail. Section V discusses experimental settings and results. The last section concludes the paper and gives directions for future work.

## II. RELATED WORKS

Recently, different Arabic SA models based on ML approaches have been proposed with various features and classifier algorithms for social media and microblogging services. Most of the proposed approaches represent documents based on the BoW model, and try to extend word features with different features such as n-grams (e.g., bi-gram, tri-gram), stems and POS tags.

Shoukry and Rafea proposed a sentiment classification for Arabic tweets [12]. They investigated using different sets of n-gram features with SVM and NB classifiers. Duwairi and Qarqaz in [13] built a SA model for Arabic Twitter and Facebook comments. In their model, the texts were represented as a set of word bi-gram features. They also investigated the effect of using term frequency (TF) and term frequency-inverse document frequency (TF-IDF) weighting schemes with SVM, NB, and K-nearest neighbours (K-NN) classifiers. Abdul-Mageed et al. in [14] presented a subjectivity and sentiment analysis system (SAMAR) based on a SVM classifier for different Arabic social media applications: Web forums, chat, Wikipedia Talk Pages, and Twitter. They studied different features including word n-grams, POS tagging, and word stems. Also, many stylistic features related to social media applications were investigated. The results showed that the classifier performance relied on the type of the dataset and features used.

Duwairi [15] proposed a SA approach for Arabic tweets written in Jordanian Arabic dialectal and Modern Standard Arabic (MSA). The researcher suggested improving the performance of the SVM and NB classifiers by transforming words in tweets from their dialect form to MSA. Hammad and Al-awadi in [16] focused on studying SA on Arabic hotel reviews collected from Twitter, Facebook, and YouTube. They employed NB, VSM, DT and back-propagation neural network (BPNN) ML classifiers with BoW, POS tag and stem features. The results showed that among the classifiers, the SVM classifier achieved the best average accuracy, followed by NB, DT and finally BPNN. Elghazaly et al. [17] evaluated the use of two classifiers, SVM and NB, on the SA of Egyptian

political election tweets. The tweets were represented using BoW features with a TF-IDF weighting scheme. The results showed that the NB classifier achieved better accuracy and a faster time than the SVM.

Despite the efforts made in previous approaches, they still suffered from a common limitation: they were semantically weak. They ignored the semantic relationship between words in the documents. Different areas of text mining in the English language, such as text clustering, topic classification and SA, have recently seen an increase in research in an attempt to cope with the BoW model's limitation. To cope with the limitation, they built semantic text representation models that incorporate semantic concepts as features using an external KB, such as WN or named entity tools.

Hotho et al. [3] are considered among the first to propose a semantic representation using WN concepts as features for clustering fields. Three representation strategies were suggested: 1. Add concepts (AddC) as extra features to the BoW model. 2. Replace words with their concepts (ReplC). 3. Use bag-of-concepts (BoC) features only. Different word sense disambiguation (WSD) strategies were followed: selecting the first concept (FstC), all concepts (AllC), and disambiguation by context. The TF-IDF weight was applied with a k-means clustering algorithm. The experiments showed that using the semantic WN concepts features were promising and outperformed the baseline BoW model. Also, Baghel and Dhir in [18] proposed a hierarchy clustering algorithm to cluster the documents based on the concepts representation. The concepts were extracted from WN using the FstC WSD strategy. The TF-IDF weighting scheme was used. The proposed approach achieved better performance than traditional approaches.

A concept-based representation approach for topics-based classification of news articles was proposed by Elberrichi et al. [4]. The proposed approach utilized WN concepts to represent documents via various representation strategies: add concepts as extra features to the BoW model, replace concepts with words, and use BoC features only. Two WSD methods were used, FstC and AllC. The classifier model applied the TF-IDF weight with cosine distance similarity. The experiments showed that using the semantic WN concepts features with AddC and FstC WSD methods outperformed the baseline BoW model.

In the SA field, Balamurali et al. [8] proposed using WN concepts features to represent texts in travel reviews datasets. Two incorporation strategies were used, AddC and BoC, with two WSD methods, manual and automatic. They used the SVM classifier and found that using the AddC strategy with the manual WSD method achieved the best performance with an accuracy of 90.20%, which increased the performance of the SA by 5.3 % over the baseline BoW.

Gautam and Yadav in [7] proposed a semantic WN synonyms analysis method for SA on a Twitter dataset. The approach relied on checking the semantic synonym similarity between words in the testing and training tweets datasets. If a synonym similarity between the words was found, words in the testing data would be replaced with their synonyms in the training data. The approach was evaluated using different ML classifiers. The results showed that the NB classifier with TF

weight obtained the superior results compared to the other classifiers used, SVM and maximum entropy (ME).

Another significant approach for SA on Twitter data was proposed by Saif et al. [19]. The approach utilized the semantic concepts, extracted from named entity tagger tools, as an additional feature into a training dataset for SA. The approach was based on the idea that specific entities and concepts tend to have a more consistent correlation with positive or negative sentiments. Knowing these correlations helps determine the sentiment of semantically relevant entities, even if those entities never appeared in the training set. They used the TF weight schema and NB classifier. The proposed approach outperformed the baseline feature BoW with POS.

### III. ARABIC TWITTER SENTIMENT ANALYSIS MODEL

A SA model for Arabic Twitter data based on the ML approach was developed. The overall architecture of the ATSA model consisted of two main phases, training and testing. In the training phase, the classifier needed to learn from a set of labeled tweets. It was then used to classify unlabeled tweets in the testing phase. Each phase consisted of the following steps: text preprocessing, features extraction, and classification. The general process of the ATSA model is illustrated in Fig. 1. First, the tweets datasets needed to be collected and annotated. After that, the tweets were preprocessed to eliminate the noise. Then, the features representation model was constructed. This step is critical because the type of extracted features and the manner in which they are built influences the performance of the ML classifier. Two different types of features were extracted, BoW and semantic concepts features, which were used to build the texts representation models. Finally, the ML classifier is trained and evaluated on unlabeled data. This section discusses the ATSA steps in more detail.

#### A. Text Preprocessing

Text preprocessing is an essential step in microblog data to clean the input tweets and eliminate noise and unnecessary data [20, 21]. Preprocessing consisted of the following steps: adding tags, data cleaning, normalization, tokenization, and stop words removal.

##### 1) Adding Tags

In Twitter, people express their sentiments using different emoticon symbols. Thus, the emoticon symbols in tweets require special handling. Furthermore, some of the punctuation marks, such as exclamation mark ("!") and question mark ("?"), are related to people's emotions [22]. In this step, emoticons symbols are replaced with their corresponding meaningful word tags that represent their sentiment. Examples of the used emoticons are displayed in Table 1.

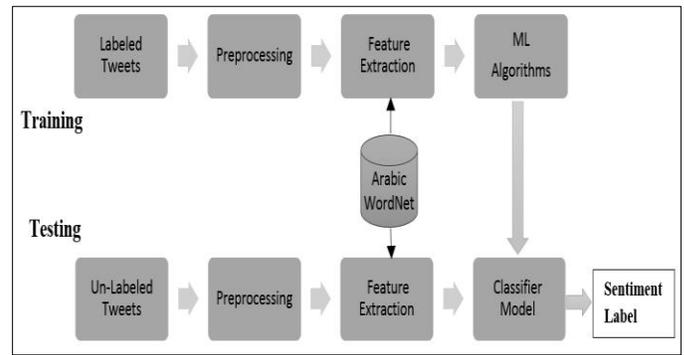


Fig. 1. Main ATSA Phases

##### 2) Data Cleaning

Data cleaning is a critical task for dealing with the noisy nature of Twitter data. This step consisted of removing items from tweets that do not include any sentiments. As such, the following items were removed: URLs, re-tweet (RT) entities, usernames, numbers, single Arabic letters, non-letter characters (e.g., + = % \$), and punctuation marks except question marks and exclamation marks (e.g., . , : "" ; ').

##### 3) Normalization

The normalization task is important in order to produce consistent word forms. Normalization for the Arabic text consisted of the following steps:

- Stripping diacritics: e.g. " العَرَبِيَّة " to " العربية ".
- Stripping lengthening (Tatweel): e.g. " العربية " to " العربية "
- Removing "ال" from the beginning of words: e.g. " العربية " will be " عربية "
- Replacing the letter "ة" with "ه"
- Replacing the letter "ى" with "ي"
- Replacing the letters "أ-إ-آ" with "ا"
- Normalizing repeated letters: e.g. " سعاااaa " to "سعادة"

##### 4) Tokenization

In this step, the tweet text was split into a sequence of tokens where each token represents a single word based on whitespaces.

##### 5) Stop Words Removal

Removing stop words is a common step in text preprocessing. Stop words (such as from, in and of) are very common words that are frequently repeated in the dataset and

do not provide any useful information to the text analysis. Removing these words allows the focus to be on the more important words and helps in dimension reduction. A list<sup>1</sup> of stop words was extended with many Arabic informal dialect words such as "عشان, كذا". The stop words from the list, except for negations, were removed.

TABLE I. EXAMPLES OF THE EMOTICONS WITH THEIR TAGS

Symbols	The Tags
“:)”, “:-)”, “:)”	HAPPY
“:(”, “-:”, “:(”	SAD
!	Exclamation
?	Question

### B. Features Extraction

Supervised ML algorithms require an appropriate representation of the documents as a features vector. The vast majority of ML approaches use the VSM [2], where each document is represented as a weighted features vector.

Different text representation models were created for tweets based on two extracted features: BoW and semantic concepts. The features need to be weighted using the term frequency-inverse document frequency (TF-IDF) [20] weighting scheme. This scheme helps reduce the weight of the features that appear in multiples dataset documents. It is defined as:

$$TF-IDF(f_n, di)=TF(f_n, di) \cdot IDF(f_n) \tag{1}$$

where  $TF(f_n, di)$  is the frequency of the feature  $f_n$ ,  $IDF(f_n)$  is defined as:

$$IDF(f_n)=\log \frac{|D|}{DF(f_n)} \tag{2}$$

where  $DF(f_n)$  refers to the number of documents in  $D$  that include the feature  $f_n$ . The  $|D|$  is the total number of documents in the dataset.

#### 1) The Bag-of-Words Representation

The BoW model used in most text mining applications has been shown to be quite effective in the SA field. To build the feature vectors, it considers the words as basic informative aspects of the texts. It consists of distinct words that appear in the dataset after preprocessing the tweets. Rather than depending on word features only, emoticon symbols are used as extra features with the BoW model to indicate the sentiment of the Arabic texts.

#### 2) Concepts Representation

Representing tweets with BoW models neglects the semantic associations between words. As a result, synonymous words that appear in two tweets are represented as different independent features, and the model would not detect any related features between the tweets. This work proposed representing the tweets in their semantic space by incorporating semantic concepts to the tweets' features space. This helps classify the sentiment of tweets that did not mention any words found in the training dataset, but did contain similar synonymous words. The concepts representation approach is described in detail in Section IV.

<sup>1</sup> Available From: <https://code.google.com/p/stop-words/>

### C. Classification

In this step, the resulted representation is supplied to the ML classification algorithm to build and learn a classifier model from training labeled tweets that can predict the sentiment label of new unlabeled tweets. Various supervised ML classifiers have been applied in previous research work on SA. The ATSA model was evaluated using the most common algorithms NB and SVM.

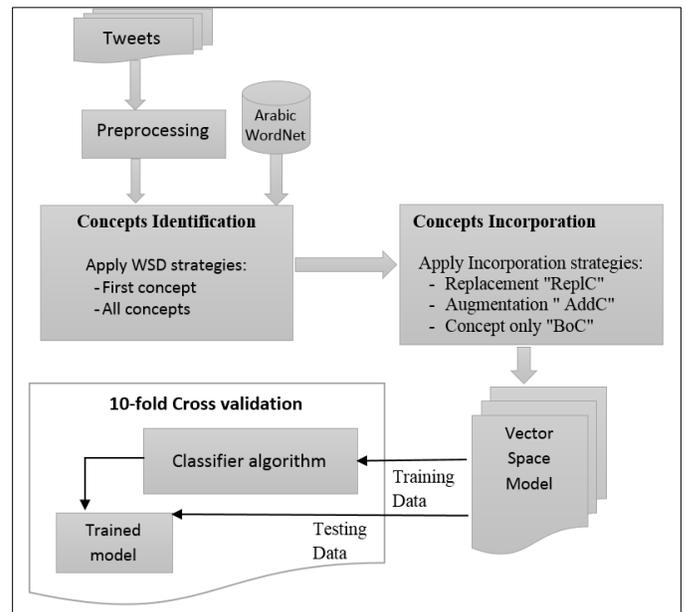


Fig. 2. Semantic ATSA architecture

## IV. CONCEPTS REPRESENTATION APPROACH

Arabic WordNet (AWN) [23] is a lexical and semantic recourse of the Arabic language based on the English Princeton WordNet. It semantically groups words together into concepts based on their meaning. A concept, also named a synset, is a basic object in the WordNet to express a set of synonym words that share at least one sense.

The proposed concepts representation approach depends on extracting and employing semantic concepts features utilizing AWN, as shown in fig 2. To develop the concepts representation model, two main steps are required:

- Identify the concepts features.
- Incorporate the concepts using different strategies.

#### A. Concepts Identification

The target of this task is to identify the concepts in tweets by utilizing the AWN ontology. To extract the concepts, a number of steps are performed. First, words in each tweet are mapped to their concepts. The WordNet returns an ordered list of all related concepts. They are ordered from most appropriate to least appropriate. Then, for words that have many concepts, it is important to select the most appropriate meaning using the WSD strategy. While building an advanced WSD approach is beyond the scope of this research, the research concentrated on simply determining whether the WSD strategy was needed to produce a good performance. The simplest WSD strategies that

were used in previous works [3, 4, 7, 8, and 18] were applied; they are first concept (FstC) and all concepts (AllC) strategies.

- First Concept (FstC): This strategy selects the first concept from the returned list as the disambiguation method.
- All Concepts (AllC): It is considered a basic strategy that selects all concepts of the word from the returned concepts list.

### B. Concepts Incorporation

In this step, the extracted semantic concepts from tweets are incorporated as extra features to represent the tweets. It was proposed to use different incorporation strategies (augmentation "AddC", replacement "ReplC", and concept only "BoC") which have been previously developed in [3, 4, and 8] for English text mining applications.

#### 1) Augmentation "AddC"

This strategy augments the identified concepts in the tweets into the BoW model as additional features with their corresponding words. By using the "AddC" strategy, the tweets are represented by all the extracted concepts and all the tweet's words. In this strategy, the size of the features is enlarged by the semantic concepts, and the new size is defined as  $|F'| = |F| + |C|$ . Where  $|F'|$  is the total number of features,  $|F|$  is the primary feature size, and  $|C|$  is the number of semantic concepts associated with the words.

#### 2) Replacement "ReplC"

This strategy replaces all words with their mapped concepts identified in the tweets. By using the "ReplC" strategy, tweets represented by concepts and words which have no map concepts in AWN. This strategy helps in reducing the features space, where the new size is defined as  $|F'| = |F| - |W_c| + |C|$ , where  $|W_c|$  is the total number of individual words that are substituted by concepts.

#### 3) Concept Only "BoC"

In this strategy, tweets are represented by their extracted concept only without any of their words. By using the BoC strategy, the size of the feature space is the same as the extracted semantic concepts  $|F'| = |C|$ .

## V. EXPERIMENTS AND EVALUATION

In this work, different tweet representation approaches were experimented to determine the best approach that improved the performance of ATSA model. The proposed semantic concepts representation against the basic BoW features was compared. Moreover, the variations of the semantic representations that resulted from applying the BoC only, AddC, and ReplC strategies with AllC and FstC WSD methods were evaluated. All of the experiments were

conducted with two ML classifiers, SVM and NB, using RapidMiner<sup>2</sup>, which is a popular data mining tool.

### A. Arabic Twitter Dataset

An Arabic Twitter corpus was built for SA by collecting tweets regarding people's sentiments in different domains (politics, sports, social and companies). To automate the process, a python script was implemented to collect data from the Twitter API<sup>3</sup> using the Tweepy<sup>4</sup> library. The collection process was based on certain hashtags that represent important events or topics for each domain.

After collecting the tweets, each tweet was manually annotated with a sentiment class label as positive or negative. In this step, two human annotators were asked to read the tweets and assign sentiment labels to them. Most of the time, they agreed about the sentiment label. When they disagreed, a another human annotator was asked to determine the final label. The final dataset consisted of about 826 tweet text documents consisting of 413 positive and 413 negative tweets.

### B. Evaluation Method and Performance Measurements

The performance of the developed approach was evaluated using F-measure. It is the harmonic mean of precision and recall. Precision and recall are two standard evaluation metrics widely used to evaluate the effectiveness of classification algorithms on a given category [24, 25].

The Precision (P), is the number of correctly classified positive tweets divided by the number of tweets labeled as positive by the system. It is defines as:

$$P = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \quad (3)$$

The Recall (R), is the number of correctly classified positive tweets divided by the number of positive tweets in the dataset. It defines as:

$$R = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \quad (4)$$

Given P and R, the F-measure is defined as:

$$F = 2 \cdot \frac{P * R}{P + R} \quad (5)$$

To ensure reliable results, all of the experiments were conducted using a ten-fold cross validation method [24, 25].

TABLE II. F-MEASURE VALUE OF NB AND SVM CLASSIFIER WITH ALL REPRESENTATION APPROACHES

features	NB		SVM	
	AllC	FstC	AllC	FstC
BoW	85.99		91.15	
Concepts only	85.96	89.14	87.04	93.7
Concepts (AddC)	89.72	91.77	92.59	95.63
Concepts (ReplC)	86.79	88.93	90.87	95.11

<sup>2</sup> <https://rapidminer.com>

<sup>3</sup> <https://dev.twitter.com>

<sup>4</sup> <http://www.tweepy.org>

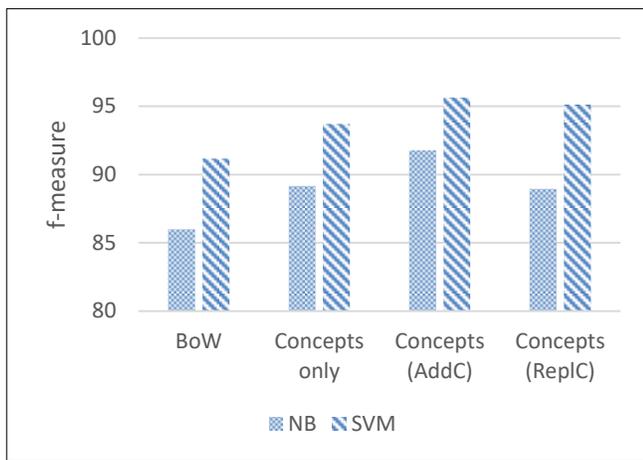


Fig. 3. The values of the F-measure for NB and SVM classifier with different representation approach and FstC WSD

### sResults

Table 2 displays the results of different tweets representation with the NB and SVM ML classifiers. The semantic representation model using AWN concepts features was found to help improve the accuracy of the ATSA model. That is because the representation of the text was enriched with the semantic concepts feature which helped preserve the semantic relations between the words, such as synonyms, and produce more common concepts features that identify the related sentiment class. For example, “فرح” (happy) and “مسرور” (glad) are synonymous words and both carry positive sentiments, the synonym relation between the words can be preserved only if the words were treated as concepts, not just as independent words.

As shown in fig 3, the performance of all of the classifiers improved with all of the proposed concepts representations. Furthermore, using the SVM classifier was found to outperform the NB classifier in almost all representation models. The highest F-measure value reached 95.63% when using the AddC concepts representation with the FstC WSD methods and SVM classifier.

Also, from the results, it is clear that using the AddC incorporation strategy provides the best performance over all concepts representation approaches. The BoC only representation discards the words that do not appear in the AWN. So, in this case, it may lose some of the distinctive word features which represent the sentiment class.

Moreover, regarding the effect of WSD, it is obvious from the experimental results that using simple FstC WSD outperformed the AIIc method with almost all concepts representation, as illustrated in fig 4. Thus, using all the concepts could produce some noise data and mislead the sentiment classification.

### VI. CONCLUSION AND FUTURE WORKS

In this work, the effect of using semantic AWN concepts features to represent tweets on the proposed ATSA model was demonstrated. Various approaches were proposed for building the concepts representation model using BoC only or

combining the BoW with the concepts following two strategies: concept augmentation and concept replacement.

The experiments showed that using concepts features outperforms the baseline BoW model and opens great opportunities to build a robust SA model for Arabic tweets. Furthermore, among all of the approaches, augmentation concepts representations with the FstC methods achieved the best accuracy.

For future, the researchers plan to examine the semantic concepts representation model on larger datasets. Also, the conducted experiments proved that using a simple WSD method had a good effect on the concepts representation. Thus, developing more advanced WSD methods is critical for the Arabic language. Moreover, the researchers suggest developing an approach for extracting the concepts features from Wikipedia and using them to extend the representation of Twitter data.

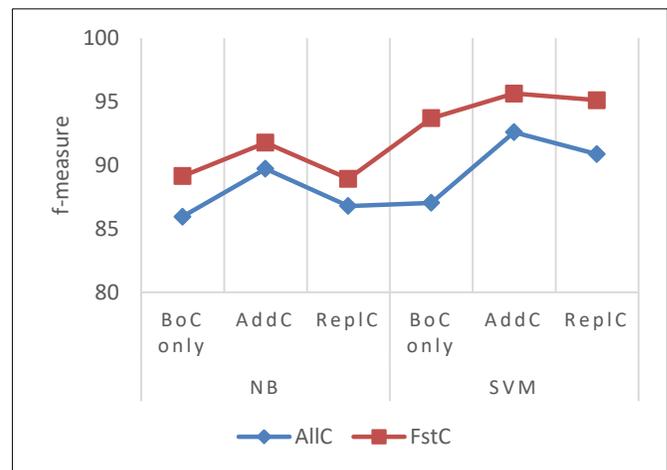


Fig. 4. Comparison of WSD Methods on the Concepts representation with Different Classifier

### REFERENCES

- [1] B. Pang and L. Lee, "Opinion Mining and Sentiment Analysis," Found. Trends Inf. Retr., vol. 2, pp. 1-135, 2008.
- [2] G. Salton, A. Wong, and C.-S. Yang, "A vector space model for automatic indexing," Communications of the ACM, vol. 18, pp. 613-620, 1975.
- [3] A. Hotho, S. Staab, and G. Stumme, "Wordnet improves Text Document Clustering," 2003.
- [4] Z. Elberrichi, A. Rahmoun, and M. A. Bentaallah, "Using WordNet for Text Categorization," Int. Arab J. Inf. Technol., vol. 5, pp. 16-24, 2008.
- [5] Z. Elberrichi and K. Abidi, "Arabic text categorization: a comparative study of different representation modes," Int. Arab J. Inf. Technol., vol. 9, pp. 465-470, 2012.
- [6] S. A. Yousif, V. W. Samawi, I. Elkabani, and R. Zantout, "The Effect of Combining Different Semantic Relations on Arabic Text Classification."
- [7] G. Gautam and D. Yadav, "Sentiment analysis of twitter data using machine learning approaches and semantic analysis," in Contemporary Computing (IC3), 2014 Seventh International Conference on, 2014, pp. 437-442.
- [8] A. Balamurali, A. Joshi, and P. Bhattacharyya, "Robust sense-based sentiment classification," in Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis, 2011, pp. 132-138.

- [9] Internet World Stats, "Top Ten Internet languages - world Internet statistics," 2016. [Online]. Available: <http://www.internetworldstats.com/stats7.htm>. Accessed: Oct. 29, 2016
- [10] M. K. Saad and W. Ashour, "Arabic morphological tools for text mining," *Corpora*, vol. 18, p. 19, 2010.
- [11] A. Farghaly and K. Shaalan, "Arabic natural language processing: Challenges and solutions," *ACM Transactions on Asian Language Information Processing (TALIP)*, vol. 8, p. 14, 2009.
- [12] A. Shoukry and A. Rafea, "Preprocessing Egyptian dialect tweets for sentiment mining," in *The Fourth Workshop on Computational Approaches to Arabic Script-based Languages*, 2012, p. 47.
- [13] R. M. Duwairi and I. Qarqaz, "Arabic Sentiment Analysis using Supervised Classification," in *Future Internet of Things and Cloud (FiCloud)*, 2014 International Conference on, 2014, pp. 579-583.
- [14] M. Abdul-Mageed, S. Kuebler, and M. Diab, "SAMAR: A system for subjectivity and sentiment analysis of social media Arabic," in *3rd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA)*, ICC Jeju, Republic of Korea, 2012.
- [15] R. M. Duwairi, "Sentiment analysis for dialectical Arabic," in *Information and Communication Systems (ICICS)*, 2015 6th International Conference on, 2015, pp. 166-170.
- [16] M. Hammad and M. Al-awadi, "Sentiment Analysis for Arabic Reviews in Social Networks Using Machine Learning," in *Information Technology: New Generations*, ed: Springer, 2016, pp. 131-139.
- [17] T. Elghazaly, A. Mahmoud, and H. A. Hefny, "Political Sentiment Analysis Using Twitter Data," in *Proceedings of the International Conference on Internet of things and Cloud Computing*, 2016, p. 11.
- [18] R. Baghel and R. Dhir, "A Frequent Concepts Based Document Clustering Algorithm," *International Journal of Computer Applications*, vol. 4, pp. 6-12, 2010.
- [19] H. Saif, Y. He, and H. Alani, "Semantic sentiment analysis of twitter," in *The Semantic Web-ISWC 2012*, ed: Springer, 2012, pp. 508-524.
- [20] C. C. Aggarwal and C. Zhai, *Mining text data: Springer Science & Business Media*, 2012.
- [21] M. Ikonomakis, S. Kotsiantis, and V. Tampakas, "Text classification using machine learning techniques," *WSEAS transactions on computers*, vol. 4, pp. 966-974, 2005.
- [22] C. Quan and F. Ren, "Construction of a blog emotion corpus for Chinese emotional expression analysis," in *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 3-Volume 3*, 2009, pp. 1446-1454.
- [23] W. Black, S. Elkateb, H. Rodriguez, M. Alkhalifa, P. Vossen, A. Pease, et al., "Introducing the Arabic wordnet project," in *Proceedings of the third international WordNet conference*, 2006, pp. 295-300.
- [24] F. Sebastiani, "Machine learning in automated text categorization," *ACM Comput. Surv.*, vol. 34, pp. 1-47, 2002.
- [25] G. Forman and M. Scholz, "Apples-to-apples in cross-validation studies: pitfalls in classifier performance measurement," *SIGKDD Explor. Newsl.*, vol. 12, pp. 49-57, 2010.