New Avenues in Arabic Sentiment Analysis

Manal Mustafa	A.Shakour AlSamahi	Alaa Hamouda
Assistant Lecturer at the Engineering faculty of Azhar University. Her research interests include software engineering, text mining. Manal has an MS in software engineering from Azhar University. Contact her at manalmustafa10@yahoo.com	Professor at the Engineering faculty of Azhar University. His research interests include data mining, and bioinformatics, software engineering. AlSamahi has Ph.D and MS in Microprocessor from Walze University.U.K amsamahy@windowslive.com	Associate Professor at the Engineering faculty of Azhar University. His research interests include data mining, and bioinformatics, software engineering. Hamouda has Ph.D and MS from Azhar University alaa_ham@gega.net

Abstract— This paper explores a method that analyzes Arabic text lexically, morphologically and semantically. The highly agglutinative nature of Arabic diminishes the effectiveness of conventional Bag of Words (BoW) which considered insufficient to form a representative vector for large scale social media content as it ignores possible relations between terms. The proposed work overcomes this limitation by incorporating different feature sets and performing cascaded analysis that fundamentally contains lexical analysis, morphological analysis, and semantic analysis. ICA is used to handle Arabic morphological pluralism. AWN semantically is exploited to extract generic and semantic relations for the lexical units over all the dataset. Moreover, specific feature extraction components are integrated to account for the linguistic characteristics of Arabic. Finally, we can leverage from standard social media features such as emoticons and smileys. So, a system for automatic Emotion Detection (ED) and mood recognitions was built to provide further sentiment insight and classification power. The optimal feature combination for each of the different emotions was determined using a combination of Machine Learning (ML) and rule-based methods. Experimentally, the results revealed that incorporation of multifaceted analysis is superior to classical BoW representation, in terms of feature reduction (31% reduction percentage) and accuracy results (F- Measure was increased up to 89%).

Index Terms— Arabic, Sentiment analysis, Morphological analysis, Semantic relations, Emoticons, Social network, SVM, RF

1 Introduction

WEB 0.2 has evolved to become a source of varied kind of Information [41] [32] [9]. Social networks (e.g., Facebook, Twitter, Google+ ...etc) represent an essential part of Web 2.0, where a massive volume of textual reviews are generated by its users. People post real time messages about their opinions on a variety of topics, discuss current issues, [2]complain, collaborate, share emotions, express sentiment for products they use in daily life and interact to each other in different virtual societies [28] [15] [6]. So, it becomes a troublesome issue for a human reader to discover related sites, take out the related contents with opinions, investigate, outline, and sort out them into perfect compatible and usable structure [20]. Hence, this valuable mine of text needs to be discovered and analyzed automatically [23] [12]. Therefore, the last ten years witnessed an intense interest in Sentiment Analysis (SA) worldwide [21]. While a lot of effort has been put into Western languages mostly English, minimal experimentation and little attention have been tailored to Arabic [42] [36] [34] [24] [22] [21] [5].

Arabic language presents significant and central challenges to many NLP applications due to its highly morphological richness [26] [24] [21] [5] [4]. Arabic language is also interesting because of its history [37], the strategic importance of the Arabic region, rapid growth in terms of online users, content, and its cultural and literary heritage [22] [13]. In particular, dialectal Arabic mainly used to express customer views about different aspects of life within social media sites [30]. The main problem of colloquial Arabic is the lack of standardization [34]. Grammar and rules govern the use of Modern Standard Arabic (MSA), but colloquial Arabic lacks a grammar and rules showing how to use it [23]. The main problems are: Arabic has very rich inflectional mechanisms and is considered one of the richest languages in terms of morphology [24], the agglutinative and derivational nature [14] [10], scarcity of lexical and linguistic resources [40] [37], dialects have not standard orthographies [30] widespread of synonyms, lack of publicly and freely accessible corpora [29] [10] [5], free order and no capitalization [37].

Furthermore, with the global acceptance of social networks, Emotion Detection (ED) has become useful and more feasible. However, most of the current works mainly target the English with minimal interest for Arabic [22]. Hence, the main interest of this paper shows how to best represent lexical information; how Arabic morphological challenges and its dialect can be soundly anticipated; whether semantic concepts are useful for Arabic SA, can we build a model for investigating mood and describing emotion; what it is impact on performance. Finally, can we build a multifaceted system and design an integrated framework that automatically detects and summarizes an overall Arabic sentiment taking into account genre types of features.

Accordingly, the contribution of this study is:

(1) Arabic morphological pluralism was automatically handled by extending ICA taking into account all parts of speech and sentence contextual structure with specific focus on dialect.

(2) Introduces and implements a new set of semantic orientations for training a model to provide further sentiment insight and classification power.

(3) Building ED model for expressive emoticons, smileys and punctuation categorization that have a measurable impact on performance.

(4) Employing standard features of social networks to despite the inherently short texts typically used in these genres.

(5) Minimizing the tedious effort required through human annotation of messages.

(6) The utility of all features is evaluated and compared for sentiment classification of Arabic content.

The rest of paper is structured as follows. Section 2 surveys related work. Section 3 propose a system that claims the ability to understand natural language segments, analyze them objectively, lexically, morphologically, as well as semantically, detect emotional behavior and extract novels and rich features. Section 4 devoted for classification techniques. Experiments and results are discussed in section 5. Finally, section 6 concludes the paper and points to possible directions for future work.

2 Related Work

The most basic strategy used by researchers concentrates on the simplest of lexical features, the Bag of Words (BoW) [3] to form a vector space model (VSM) [11].Using single words as a representative feature in Text Classification (TC) has proven effective for a number of applications [25]. However, this feature lacks semantic information to classify text accurately [10]. Furthermore, BoW representation suffers from a huge feature vectors that should be carefully considered to avoid hardware limitation, software capabilities, and computational time complexity [11]. So [11] used the singular value decomposition (SVD) method to extract textual features based on latent Semantic Indexing (LSI) to classify Arabic text. A study presented by [21] investigated Logistic Regression (LR) as a new algorithm for Arabic TC Area. Chi Square (CHI) is used for Feature Selection (FS) and a Normalized Frequency (NF) is used as a weighting scheme for term representation in the VSM. A Similar comprehensive comparative study of the different tools for Arabic text preprocessing, attribute selection, reduction and classification include the work of Khorsheed and Al-Thubaity [25].

Arabic Sentiment Analysis (ASA) is hindered due to lack of resources [37]. So, [34] survey the research efforts to analyze the Arabic content in Twitter focusing on the tools and methods used to extract the sentiments for the Arabic content on Twitter. [13] presented SA for MSA and Egyptian dialect with a corpus of different types of data. They employed a number of novels and rich features that include valence shifters, negation, intensifiers, questions and supplication terms to improve the classification performance. In other work, Nizar et.al [29] considered the multi-way SA problem for Arabic reviews and coupled BoW with the most popular classifiers. Abdul-Mageed et.al [26] in their contribution, propose SAMAR system which adopts two-stage classification approach; subjectivity and SA for Arabic social media. SAMAR uses different feature sets such as lexical, morphological but did not use semantic features. A similar finding was reported by [1] who use a genetic algorithm for both English and Arabic Web forums sentiment detection on the document level. They track both syntactic and stylistic features, but do not use morphological features.

A few researches concerning ASA used Arabic WordNet (AWN) for improving classification results such as [42] and [10]. [42] estimate three different representation modes; BoW, N-Grams, and AWN as lexical and semantic resource to improve Arabic TC. [10] present semantic model-based representation of text and considered two ways of grouping synonyms: synonym's grouping using stems and synonym's grouping using roots to improve the representation and reduce the dimensionality.

In SA, emotion provides practically a promising direction for fine-grained analysis of subjective content [16] [3]. In recent years, the study of emotions has obviously proved how emotions are fundamental to human experience, influencing cognition, perception, communication and decision-making. Erik [8] reported that the basic tasks of affective computing and SA are emotion recognition and polarity detection. So, it is crucial for automated SA methods to correctly account for such graphical cues for sentiment [3]. [18] explored the microblogging features including emoticons, abbreviations and the presence of intensifiers such as all-caps and character repetitions for Twitter SA. Their results show that the best performance comes from using the n-grams together with the microblogging features and the lexicon features where words tagged with their prior polarity.

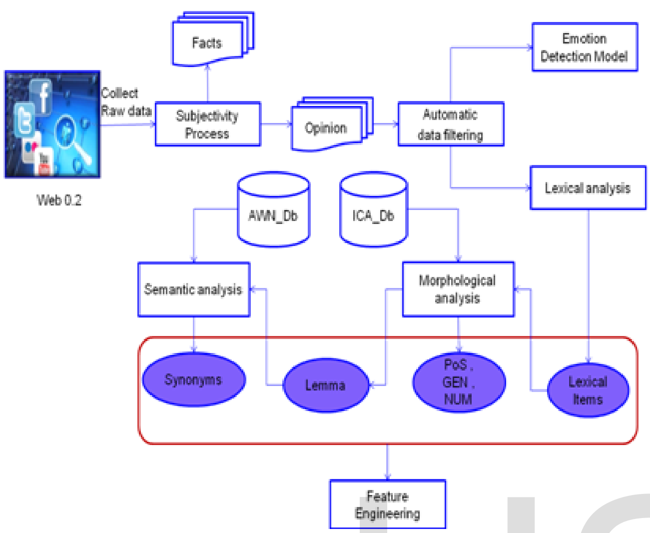
However, most of these attempts are based on statistical approaches applied on BoW and lack a common framework that can combine genre types of features. Lexical features only capture local information in the data and do not take into consideration of morphological configuration and lacks semantic dimensions as well as emotions and social networks information. Therefore, there is an urgent need to develop an intelligent and a multifaceted system which incorporates different levels of text analysis in a cascade configuration to accurately categorize Arabic social media contents. Additionally, a methodology is developed for processing of massive volumes of streaming data, as well as the automatic identification of human expressiveness within short text messages. Moreover, the suggested framework addresses different dimensions of opinions, such as subjectivity (Subjective vs. Objective), polarity (Positive vs. Negative), and efficiently describes emotional state.

ISSN 2229-5518

3 Proposed Work

The proposed method exhibits sequentially six main stages are graphically presented in Figure 1 and discussed during the subsequent sections.

Figure 1 - Proposed Framework



3.1 Subjectivity Process

Supervised learning methodology was followed where each tweet is modeled as a vector of sentiment features with the corresponding label. However, SA solutions ultimately face the challenge of separating the factual content from the subjective content describing it [9][6]. Therefore, two classification tasks are considered: subjectivity and sentiment classification [20]. In the former, tweets are classified as subjective (non-neutral) or objective (fact or neutral), and in the latter, positive and negative tweets are considered as subjective [41] [40] [31] [26]. Facts can be generalized by exploiting of special dictionaries (so-called filters), containing synonyms for neutral appraisals. The main flaw of this approach is the necessity of manual selection of terms for the filters that can extract out facts and exclude neutral concepts. Afterwards, a simple filtering method depending on some predefined emotion dictionaries had developed in order to automatically separate between opinions with emotions and opinions without emotions. The former will be handled during the ED model while, the later will be the input to the lexical analysis stage.

3.2 Lexical Analysis

Lexical analysis is the process of converting a stream of characters into sequence of symbols known as lexical items. Preprocessing is required in order to overcome the challenge of massive volume of the vector space [11]. Hence, the input stream undergoes a series of steps discussed as following:

3.2.1 Tokenization

The process of determining and classifying a clause into set of tokens by recognizing delimiters [10].

3.2.2 Normalization

Informal and non-grammatical nature of the colloquial language used in social media needs to be automatically handled. The process is needed to put a text in a consistent form and converts all the various forms of a word to a common form [13]. Consequently, a prototype is developed to remove diacritics, normalize different words [34] [14], remove repeated letters (if the letter repeated more than 2 times), handle franco words and convert these word to its Arabic equivalent.

3.2.3 Stemming

Light stemming is employed to remove common affixes from words and return an inflected or derived word to its stem.

3.2.4 Stop words filtering

It is also useful to ignore very common words of the messages that do not provide discriminative meaning [14] [11]. Stop words occur frequently in most of the documents in a given collection, usually, have on contribution to the semantics of the context, can't be used as index terms, most of those words are irrelevant to the categorization task, can be dropped with no harm to the classifier performance, and never form a full sentence when used alone. In addition, a new dialectical stop words list is identified {e.g., وعشان وبعدين ’وبعديها, وعلشان’ , ولسا, وكمان, عليك فينيا, واحنا, فينو’ فينا, ولسة}. Consequently, the main advantage of the pre-processing cycle is to reduce the number of terms in the corpus so as to reduce the computational and storage requirements of ML algorithms.

3.3 Morphological Analysis (MA)

The most basic type of linguistic analysis because it forms the essential foundation for further types of analysis (such as syntactic parsing and semantic field annotation). The aim of morphological analysis of corpora is not only to assign each item in the text to its identical Lemma, but also to indicate a set of morphological concepts such as a code indicating its appropriate Part of Speech (PoS) tagging, gender, numbers, person, voice etc… Actually, there are many morphological analyzers for Arabic; Buckwalter Arabic Morphological Analyzer (BAMA) and Xerox Arabic Finite State Morphology represent the best known, well documented, morphological analyzers for MSA. Yet there are significant problems with both analyzers in design as well as coverage that increase the ambiguity rate. Some of these problems are handled in International Corpus of Arabic.

3.3.1 International Corpus of Arabic

Al-Nasray et.al [35] present description and technical design of the International Corpus of Arabic (ICA). ICA has been analyzed by Bibliotheca Alexandrina Morphological Analysis Enhancer (BAMAE). In this study, ICA is employed In order to diminish data sparseness and to tag every lexical unit with the corresponding different morphological features. Therefore, ICA is arranged in SQL_Db in order to provide a tool capable of performing both a lemmatizer and a tagger efficiently. A lemmatizer automatically reduces all inflectional and variant_forms of a word to their lemma that can be substantially used in word indexes, dictionaries and concordances. On the other hand, a tagger efficiently annotates every word with the corresponding PoS. The following section highlights the morphological characteristics associated with every word.

3.3.2 Investigating Morphological Characteristics

Lemma (LEM): where the words are reduced to their lemma forms, (citation forms) instead of using terms in their orthographic form. For verbs, this is the 3rd person masculine singular perfective form and for nouns, this corresponds to the singular default form (typically masculine).
PoS tagging: Since we use only the base forms of words, the question arises whether we lose meaningful morphological information and consequently whether we could represent this information as PoS tags instead. So, PoS is used as basic form of Word Sense Disambiguation (WSD).

In addition, Arabic language exhibits two genders: masculine and feminine and three number classes: singular, dual, and plural. Some researchers suggest that there is a relationship between gender (GEN) and sentiment expression. So, ICA_Db is extended to provide three gender and number types of the underlying feature set. The three gender types are masculine, female and unknown corresponding to the set {MCL, FEM, UNKNOWN) while the three number types are singular, dual and plural corresponding to the set {SNG, DUA, PLR}. Some examples are shown in Table 1.

Table 1 – Morphological Properties of Lexical Analysis

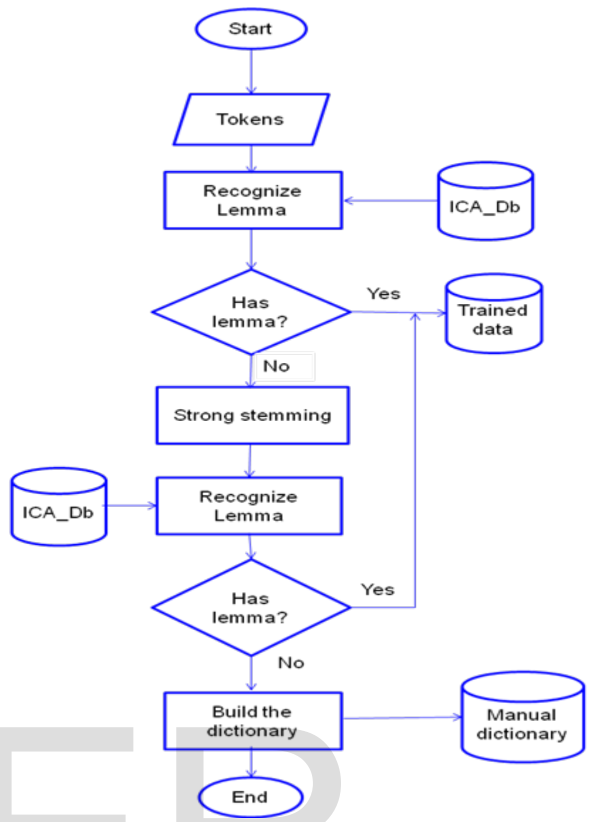
Lexical Item	Lemma	PoS	GEN	NUM
استغلوا	استغل	MCL	VER	PLR
يستغل	استغل	MCL	VER	SNG
اشتغلت	اشتغل	FEM	VER	SNG
تشتغلوا	اشتغل	MCL	VER	PLR
تشتغل	اشتغل	UNKNOWN	VER	PLR

This data model has been implemented in a SQL database. The database will be a deliverable of the project, and will be distributed freely to the community.

3.3.3 Enhancing ICA performance

Practically, the agglutinative nature of Arabic prevents ICA_DB to deal efficiently with some attributes like (ويعرف- انتصاره- فانسحب). However, a strategy was developed to overcome this problem and converting more tokens to their Lemma. The method undergoes three subsequent stages are represented in Figure 2 and discussed as follows:

Figure 2- ICA_Db and its different stages



Stage 1 simply connects to ICA_Db and finds Lemma for the main lexical items. The original form of a term is returned if it is unknown to the lemmatizer. Stage2 efficiently deals with those morphemes that have not known for ICA_Db. The agglutinative nature of Arabic language contributes this limitation which prevents lemmatizer to investigate their lemma directly. So, a strategy was developed to overcome this limitation. More aggressive stemming is done to deal with the items that have prefixes such as (ال -وال– عال -لل- بال- فال -). Such prefixes naturally connected to names and can be removed efficiently. Once strong stemming has been done, stage 3 finally builds a manual dictionary for retrieving variants and different morphemes that have no Lemma . The proposed dictionary is performed by carrying out a thesaurus lookup for every word that appears in a document. If that word matches a cluster (list) in the thesaurus then that word is replaced by that list's representative. The underlying dictionary is built according to the following manners:

1 Spelling correction was performed.
2 Making some normalization (e.g., محبتنا – محبوبا - حبي - يحبه- محبيه- محبيك)
3 Applying an extension of lexicon to include more words not restricted to MSA. As a result, some dialect words have been converted to their corresponding MSA (e.g., يشوفوني -أشوفك أشوفه –ويشوف)

4 Some franco words have been normalize to one form (e.g., توتيه- تويتات- تويتري تويتاتك- وتويته -يريتويت).

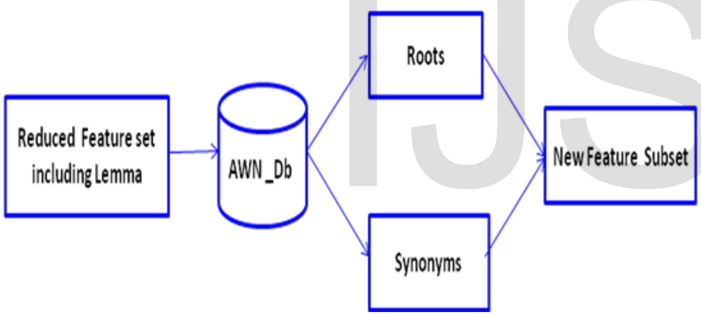
3.4 Semantic analysis using Arabic WordNet

Arabic WordNet (AWN) [42] [27] is a lexical database for MSA based on the universally accepted Princeton WordNet (PWN). AWN is extended to define semantic orientations of lexical items taking into account sentence contextual structure. Individuals frequently utilize distinctive words or phrases to depict the same feature. For example a list of key phrases “مجتهد”, “متفوق, “متميز” differs lexically but is equivalent in concept. Classifiers cannot deal with such words as correlated words that provide similar semantic interpretations. Bing Liu [6] suggests grouping those synonyms that express the same feature.

3.4.1 Synonyms grouping thesaurus

AWN is only intended for MSA and don’t take into account dialectical Arabic. However, this obstacle was overcome within the previous stage where colloquial items substantially are converted into their identical MSA. So, by using AWN synonyms relations we can choose the most prominent lexical form of a concept, and suppress other redundant phrases.

Figure 3 – Semantic Analysis using AWN

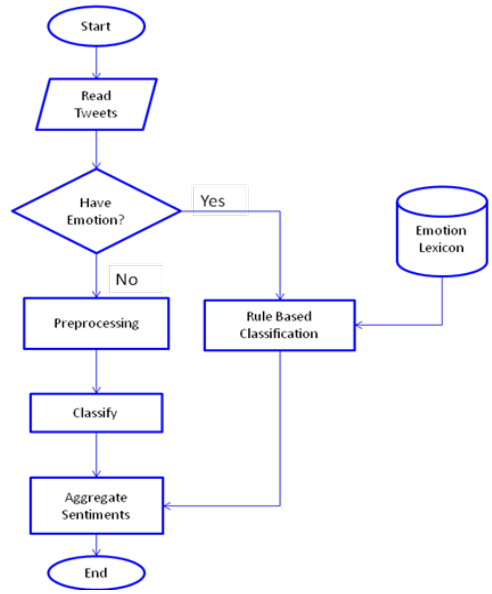


Therefore, the model aggregates synonymous words into one clusters and each cluster is represented by a single word; the one that is commonly used in that context. In addition, we can leverage from AWN roots to find some roots of the underlying features that are not computed within ICA as shown in Figure 3. By applying that cascaded analysis, size of the space vectors was fundamentally reduced up to 30%.

3.5 Emotions Detection (ED)

Supervised classification techniques were used to predict the polarities of different social media messages. However, several challenges encountered for these techniques which require labeling the data that is often expensive and time consuming [31] [7]. Since Web discourse and specifically social networks are highly emotion-related content, some heuristic rules could be defined in order to acquire merits of the presence of emoticons. The use of emoticons assumes that they could be associated with positive and negative polarities regarding the subject mentioned in the tweet [12].

Figure 4- Emotion Handling



3.5.1 Projection phase

During this phase, each text segment is checked for thepresence of emoticons as indicated in Figure 4. A number of frequently adopted emotions labels had used to construct an emotion lexicon. Emotion-oriented lexical resources provide a list of words or expressions marked according to different emotion states. An emotion-oriented method should classify the message to a special meaningful emotional category such as sadness, joy, surprise, among others. All these items are finally mapped to positive or negative category.

3.5.2 Rules generation phase

Some heuristics rules are defined in a manner that accurately classifies tweets as positive or negative according to type of emotions. However, due to the usage of mixed left-to-right and right-to-left text orientation, Arabic smileys and sad emoticons are often mistakenly interchanged, leading to contradictory sentiments within the tweet. Analysis of emotional and graphical cues of social networks yields the following:

1 Some emotions come purely with positive instances (e.g., Thumbs_ Up, heart,) as in Table 2.

Table 2– Positive Emotions Examples

Symbols	Emotion Synset	Graphical Emoticons	Sentiment
(y)	Thumbs up		Positive
♥	Heart	♥	Positive
:-*	Kiss		Positive

2 Some emotions come purely with negative instances (e.g., Devil, Thumbs_Down) as indicated in Table 3.

Table 3 – Negative Emotions Exampls

Symbols		Emotion Synset	Graphical Emoticons	Sentiment
(n)		Thumbs down		Negative
3:)	3:-)	Devil		Negative

Although there are cases where this basic assumption holds and relation between the emoticon and the tweet subject is not clear. Hence, the use of emoticons as tweet's labels can introduce noise. However, some researches counterweighted this drawback by using large amount of data that can easily be labeled [12]. In this study, a set of new hand-coded rules would defined for addressing the appearance of compound aspects consists of more than one emotion and also cover those aspects that tend to appear multiple times in a single sentence or overlap between contradict instances. Disjunctive Normal Form (DNF) is used for that purpose where the conditions are on the emotion presence. The left hand side of this formula expresses a condition on the feature set while the right hand side is the class label [39].

if (DNF formula) then (category c)

Finally, a combination of the emotional symbols with all verbal information will be taken into consideration during classification task.

3.6 Feature Engineering and Weighting Scheme

Feature extraction is a crucial SA task and is mostly ML based. A good feature vector can foresee how successful the results of the classifier will be. The extracted feature sets are indicated by the following taxonomy:

3.6.1 Structural Features

Those features represent vocabulary richness measure like average number of words per sentence, as long sentence can heavily affect the classification accuracy. Word-length distributions, total number of short words (i.e., ones less than 3 letters), character-level lexical features like average number of character for each message, special characters, letters, and presence of requoted contents.

3.6.2 Social Media Specific Features

Social media SA can be a bit challenging as individuals generally don't pay any heed to the spellings and deliberately modify the spellings of the words and use short forms whenever required. This makes it difficult to determine the overall polarity of the segments. Analysis of this speech reveals that the persons used varied repetition strategies, including anaphora, antithesis, emotions, abbreviations, intensifiers, poor spelling, poor punctuation, and poor grammar. Hence, all links were replaced with the token "URL", "HashTag" is included as a binary token to indicate if there is a hash tag contained in the sentence, all mentions of usernames (which are denoted by the @ symbol) were replaced with the token "UserName", "PosEmo" and "NegEmo" tokens are exploited to check the presence of positive or negative emotion respectively according to some predefined emotion dictionaries, normalizing all numerical digits to the token "NUM", and an "Intensifiers" token is included to indicate if a sentence contains this contextual property.

3.6.3 Negation and Punctuation Features

Negation in the tweet was experimentally considered by compiling a list of negation particles (around 30 words in Arabic) found in the tweets and checking whether a tweet contains a negation articles or not. The list contains negation terms in MAS (e.g.,لا - غير – ليس) as well as Egyptian dialect such as (e.g., مافيش- مالهاش- ماعنديش). For punctuation exclamation marks (!) and question marks (?) also carry some opinion and may be indicative of sentiment. In general, '!' is used when we have to emphasis on a positive word and '?' is used to highlight the state of confusion or disagreement and usually express negative sentiment. Hence, a binary feature used as an indicative for a segment that contains wondered expression or questionable term. Finally, the proposed approach efficiently incorporates the entire extracted new feature set into the ML classification approach as an additional feature.

4 Classification

Once the feature vectors of all the tweets from the dataset have been extracted, they are used together with the annotated sentiment labels as input for supervised learning algorithms. Several learning algorithms can be used to fulfill this task. In this study, SVM and RF are used for categorization purpose. Finally, the resulting learned function can be used to automatically infer the sentiment label regarding an unseen tweet [12].

5 Experiments and Results

5.1 Dataset Characteristics

The test set was crawled by searching Twitter API [12] specifically about famous persons. The original dataset consists of about 20000 messages. 15304 are considered as facts, news, duplicated and out of scope tweets. Actually, 4696 tweets are considered for training with the same number of positive and negative tweets. A balanced dataset is exploited to avoid the creation of models biased toward a specific class. The underlying tweets were tokenized into 11089 unique attribute as baseline.

5.2 Experiment

The experiment is performed individually on two different stages. Firstly, dataset is tokenized and constructed into its original dimensionality (F1). Secondly, the underlying feature vector is passed through ICA_Db then AWN to generate the candidate feature subset (F2) and aggregated with the entire extracted features. For each testbed, SVM is conducted as the classifier because of its reported performance [41] [17] as it gained practically good track record for SA. In practical, various kernels were investigated with the Sequential Minimal Optimization (SMO) algorithm included in Weka data mining package. Over all experiments, it is found that linear kernels yield the best performance. Moreover, Random Forest (RF) [19] is tuned as other classification algorithm to perform SA task. RF was chosen due to its superior performance over a single DT. All experiments were performed with presence vectors. SVM and RF are used with 10 fold cross validation to classify sentiments. Accordingly, efficiency of all approaches is evaluated using standard classification metrics.

5.3 Results and Evaluation

The impact of the different features was measured in isolation and incrementally are combined them in order to evaluate the robustness and efficiency of each approach. The message annotation task by the independent coders had a Kappa (k) value of 77.94% for Arabic. Precision, Recall, and F-Measure [3] are used as classification metrics. In many cases, F-measure is simply used as a measure of accuracy because it relies on both precision and recall [33]. We use previously proposed state-of-the-art (BoW) as our baseline F1. The new defined feature subset F2 represents the proposed feature vector after appending all the extracted features. The proposed method is effective in raising the F- Measure, which increased from 86.7 to 89.0 using SVM and 88.4 using RF. The relative average classification error has been reduced by almost 2.5% in comparison to the baseline method as indicated in Table 4.

Table 4- Precision, recall and F-measure rates for all modules

Classifier	Feature set	# Feature	Precision%	Recall%	F-Measure%	Feature reduction %
SVM	F1(Initial feature vector)	11089	86.8	86.7	86.7	0
	F2(all new features)	7640	89	89	89	31
RF	F2(all new features)	7640	88.3	88.2	88.4 (# trees 160)	

Additionally, RF is experimented with different number of trees to beget the highest accuracy as shown in Table 5 .

Table 5– RF Results

# Trees	F-Measure	Out of Bag Error (OOB)
10	87.2	0.1625
20	87.8	0.1371
30	88.1	0.1278
40	88.3	0.1254
50	88.3	0.1229
150	88.3	0.1175
160	**88.4**	**0.1163**

Experimentally, it is revealed that increasing in number of trees improve performance of the random model. Growing a larger forest will improve predictive accuracy, although there are usually diminishing returns once we get up to several hundreds of trees and only increase the computational cost [38]. Consequently, F-Measure was 87.2 at 10 trees and incrementally increased up to 88.3 at 40 trees. At this point F-Measure remains stable during the range from 40 to 150 trees and starts to slightly increase at 160 trees as indicated in Figure 5.

FIGURE 5 – F-MEASURE

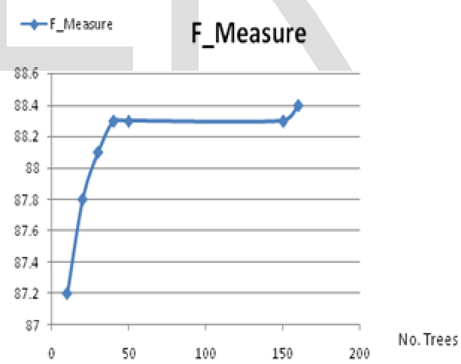
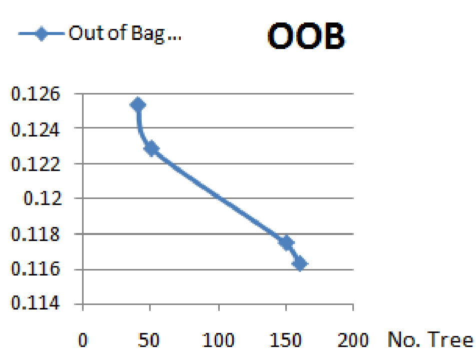


FIGURE 6– OOB ERROR



Moreover, OOB error starts at 0.126 and tends to decrease with the increasing number of trees. The best

result was at 160 trees as a tree with a low error rate is suggested to be a strong classifier [38]. So, the optimal parameters were obtained at (number of trees = 160 that minimize OOB error at 0.1163) as depicted in Figure 6.
On the other hand, F2 provides a considerable dimensionality reduction in comparison to BoW representations. That is, the presented approach avoids the sparsity problem presented by word-based feature representations for Twitter sentiment classification by reducing the feature vector with approximately 31%. And as such, F1 requires a huge amount of memory, CPU resource and extra time which perturbs the operation of the classifiers.

6 Conclusion and Future work

In this study, we propose a multifaceted and fully automated system to analyze Arabic social media contents. The features and techniques result in the creation of a SA approach geared towards classification of social media discourse sentiments with higher accuracies. Then, the study investigates the development of an ASA system using a novel rule-based approach. This approach defines some heuristic rules in a manner that accurately classifies tweets as positive or negative regardless of its contents. An illustrative example is exploited to show how the proposed method can be efficiently integrated into both SVM and RF algorithms. Therefore, a holistic and fully automated framework was built which can acquire merits of individual algorithms. A quantitative comparison among various representative vectors was efficiently performed. Such comparison indicates that, the quality of the multifaceted system outperformed that of classical individual techniques from the view point of accuracy measurements, and feature reduction percentage. The method experimentally successes to redefine the feature space with reduction percentage reach 31% for Twitter dataset. Moreover, accuracy has been increased up to 89%.
This work can be extended through incorporation of better spell correction mechanisms (may be at phonetic level) and building WSD modules to improve performance on highly ambiguous words. We can explore even richer linguistic analysis, for example, parsing, weights or scores, valence shifters and topic modeling. Also we can identify the target and entities in the tweet and the orientation of the user towards them.

REFERENCES

[1] A. Abbasi; H. Chen; A. Salem, “Sentiment Analysis in Multiple Languages: Feature Selection for Opinion Classification in Web Forums” ACM Transactions on Information Systems, Vol. 26, No. 3, Article 12, June 2008.

[2] A. Agarwal, B. Xie, I. Vovsha, O. Rambow, R. Passonneau, “Sentiment Analysis of Twitter Data”, In Proceedings of the ACL 2011 Workshop on Languages in Social Media, pp. 30–38, 2011.

[3] Alexander Hogenboom; Daniella Bal; Flavius Frasincar; Malissa Bal; Franciska de Jong; Uzay Kaymak “Exploiting Emoticons in Sentiment Analysis” 28th Annual ACM Symposium on Applied Computing, Coimbra, Portugal, pp. 703 – 710, Mar. 2013.

[4] Ashraf S. Hussein, “Visualizing document similarity using n-grams and latent semantic analysis” SAI Computing Conference (SAI), IEEE, pp. 269 – 279, Sep. 2016.

[5] Bashar Al Shboul; Mahmoud Al-Ayyoub; Yaser Jararweh, “Multi-Way Sentiment Classification of Arabic Reviews”, 6th International Conference on Information and Communication Systems (ICICS), IEEE, pp. 206 – 211, April 2015.

[6] Bing Liu “Sentiment Analysis: A Multi-Faceted Problem” IEEE Intelligent Systems, Vol.25, No.3, pp. 76 – 80, Aug. 2010.

[7] Danushka Bollegala; Tingting Mu; John Y. Goulermas “Cross-domain Sentiment Classification using Sentiment Sensitive Embeddings” IEEE Transactions on Knowledge and Data Engineering , Vol. 28, No. 2, Pages: 398 – 410, Feb. 2016

[8] Erik Cambria; “Affective Computing and Sentiment Analysis”, IEEE Intelligent Systems. Vol. 31, Issue: 2, pp. 102 - 107 March/April 2016.

[9] Erik Cambria ; Björn Schuller ,.; Yunqing Xia,.; Catherine Havasi,.; “New Avenues in Opinion Mining and Sentiment Analysis”, IEEE Intelligent Systems, Vol. 28, No.2, pp. 15-21, March /April 2013.

[10] Enas Sedki; Abdelfattah Alzaqah; Arafat Awajan, “Arabic Text Dimensionality Reduction Using Semantic Analysis”, WSEAS TRANSACTIONS on INFORMATION SCIENCE and APPLICATIONS. Vol. 12, pp. 209 – 218, 2015

[11] Fawaz S. Al-Anzi; Dia AbuZeina, “Toward an enhanced Arabic text classification using cosine similarity and Latent Semantic Indexing” Elsevier, April 2016

[12] Felipe Bravo-Marquez; Marcelo Mendoza; Barbara Poblete, “Combining Strengths, Emotions and Polarities for Boosting Twitter Sentiment Analysis”, WISDOM’13, August 11 2013, Chicago, IL, USA, ACM, 2013.

[13] Hossam S. Ibrahim; Sherif M. Abdou; Mervat Gheith, “Sentiment Analysis for Modern Standard Arabic and Colloquial”, International Journal on Natural Language Computing (IJNLC), Vol. 4, No.2, pp. 95 – 109, April 2015.

[14] Jaber Alwedyan; Wa'el Musa Hadi; Ma'an Salam; Hussein Y.Mansour, “Categorize Arabic Data Sets Using Multi-Class Classification Based on Association Rule Approach”, International Conference on Intelligent Semantic Web-Services and Applications (ISWSA’11), Amman, Jordan, ACM, April 2011.

[15] J. Bernabe-Moreno; A. Tejeda-Lorente ; C. Porcel ; H. Fujita; E. Herrera-Viedma; “CARESOME: A system to enrich marketing customers acquisition and retention campaigns using social media information” Elsevier, pp. 163–179, 2015.

[16] Jasy Liew Suet Yan; Howard R. Turtle, “Exploring Fine-Grained Emotion Detection in Tweets”, Proceedings of NAACL-HLT, Association for Computational Linguistics. San Diego, California, Pp. 73–80, June 2016.

[17] Keng-Pei Lin; Ming-Syan Chen, “On the Design and Analy-

sis of the Privacy-Preserving SVM Classifier", IEEE Transactions on Knowledge and Data Engineering, Vol. 23, No. 11, pp. 1704 – 1717, 2011.

[18] Kouloumpis, E., Wilson, T., Moore, J.: Twitter sentiment analysis: The good the bad and the omg! Proceedings of the Fifth International Association for the Advancement of Artificial Intelligence (AAAI) Conference on Weblogs and Social Media, pp. 538 – 541, 2011.

[19] Le Zhang; Ponnuthurai Nagaratnam Suganthan "Random Forests with ensemble of feature spaces" Pattern Recognition. Elsevier, Vol. 47, No. 10, pp. 3429–3437, April 2014.

[20] M.A.Jawale; D.N.Kyatanavar; A.B.Pawar; "Design of Automated Sentiment or Opinion Discovery System to Enhance Its Performance", Proc. of Int. Conf. on Advances in Information Technology and Mobile Communication, pp.48 – 58, 2013.

[21] Mayy M. Al-Tahrawi ,"Arabic Text Categorization Using Logistic Regression" I.J. Intelligent Systems Technologies and Applications (IJISA), Vol. 7, No.6, pp. 71-78, May 2015

[22] Mohammad Al-A'abed; Mahmoud Al-Ayyoub, "A Lexicon-Based Approach for Emotion Analysis of Arabic Social Media Content" Proceedings of the International Computer Sciences and Informatics Conference (ICSIC), pp. 343 – 351, 2016.

[23] Mohammed Al-Kabi; Noor M. Al-Qudah; Izzat Alsmadi, Muhammad Dabour; Heider Wahsheh "Arabic / English Sentiment Analysis: An Empirical Study" ICICS'13, April 23–25, Irbid, Jordan, ACM, 2013.

[24] Mohamed Attia; Mohsen A. A. Rashwan; Mohamed A. S. A. A. Al-Badrashiny "Fassieh; a Semi-Automatic Visual Interactive Tool for Morphological, PoS-Tags, Phonetic, and Semantic Annotation of Arabic Text Corpora" IEEE Transactions on Audio, Speech, and Language Processing · Vol. 17, No. 5, pp. 916 – 925, Aug. 2009.

[25] Mohammad S. Khorsheed; Abdulmohsen O. Al-Thubaity "Comparative evaluation of text classification techniques using a large diverse Arabic dataset" Language Resources and Evaluation, Springer, Vol. 47,No. 2, pp. 513 – 538, March, 2013.

[26] Mohammad Abdul-Mageed; Mona Diab; Sandra Kübler; "SAMAR: Subjectivity and sentiment analysis for Arabic social media" Computer Speech & Language, ACM, Volume 28, Issue 1, Pp. 20–37, Jan. 2014.

[27] Musa Alkhalifa; Horacio Rodríguez, "Automatically Extending Named Entities Coverage of Arabic WordNet using Wikipedia" International Journal on Information and Communication Technologies, Vol. 3, No. 3, pp. 20 – 36, June 2010.

[28] Nandini Burade; Ankita Kendhe " A Review on Opinion Mining through Customer Experiences" International Journal of Computer Applications(IJCA). Vol. 107, No 18, pp. 37–39, Dec. 2014.

[29] Nizar A. Ahmed, Mohammed A. Shehab, Mahmoud Al-Ayyoub and Ismail Hmeidi "Scalable Multi-Label Arabic Text Classification", 6th International Conference on Information and Communication Systems (ICICS), IEEE, pp. 212 – 217, 2015.

[30] Nora Al-Twairesh; Hend Al-Khalifa; AbdulMalik Al-Salman1, "AraSenTi: Large-Scale Twitter-Specific Arabic Sentiment Lexicons" Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics. Pp. 697–705, Berlin, Germany, Aug. 2016.

[31] Nurfadhlina Mohd Sharef; Harnani Mat Zin; Samaneh Nadali "Overview and Future [Opportunities of Sentiment Analysis Approaches for Big Data" Journal of Computer Sciences, 12 (3), pp. 153-168, 2016.

[32] Pollyanna Gonçalves; Daniel Hasan Dalip; Helen Costa; Marcos André Gonçalves; Fabrício Benevenuto, "On the Combination of "Off-The-Shelf" Sentiment Analysis Methods" Pisa, Italy, ACM, April 2016.

[33] Pollyanna Gonçalves; Matheus Araújo "Comparing and Combining Sentiment Analysis Methods". Proceeding COSN'13 of the first ACM conference on Online social networks. Boston, MA, USA, ACM, pp. 27-38, Oct. 2013.

[34] Sarah O. Alhumoud; Mawaheb I. Altuwaijri; Tarfa M. Albuhairi; Wejdan M. lohaideb, "Survey on Arabic Sentiment Analysis in Twitter", World Academy of Science, Engineering and Technology, International Journal of Social, Behavioral, Educational, Economic, Business and Industrial Engineering Vol. 9, No. 1, pp. 364 -368, 2015.

[35] Sameh Alansary; Magdy Nagi; "The International Corpus of Arabic: Compilation, Analysis and Evaluation" Proceedings of the EMNLP 2014 Workshop on Arabic Natural Language Processing (ANLP), pp. 8–17, Doha, Qatar, Oct. 2014.

[36] Sameh Alansary; Magdy Nagi; Noha Adly, "A Suite of Tools for Arabic Natural Language Processing: A UNL Approach", 1st International Conference on Communications, Signal Processing, and their Applications (ICCSPA), pp. 1 – 6, 2013.

[37] Sanjeera Siddiqui; Azza Abdel Monem; Khaled Shaalan "Towards Improving Sentiment Analysis in Arabic", Proceedings of the International Conference on Advanced Intelligent Systems and Informatics (AISI), Springer International Publishing, pp. 114-123, Oct. 2016.

[38] Thais Mayumi Oshiro; Pedro Santoro Perez; Jos´e Augusto Baranauskas "How Many Trees in a Random Forest?", MLDM 2012, Verlag Berlin Heidelberg, Springer, pp. 154–168, 2012.

[39] Vidisha M. Pradhan; Jay Vala ; Prem Balani , "A Survey on Sentiment Analysis Algorithms for Opinion Mining", International Journal of Computer Applications (IJCA) Vol. 133 – No.9, pp. 7 – 11, Jan. 2016.

[40] Walaa M.; Ahmed H.; Hoda K.; "Sentiment analysis algorithms and applications: A survey", Ain Shams Engineering Journal, ElSevier, Vol. 5, No. 4, Pp. 1093–1113, Dec. 2014.

[41] Yan Dang, Yulei Zhang, and Hsinchun Chen "A Lexicon-Enhanced Method for Sentiment Classification: An Experiment on Online Product Reviews" IEEE Intelligent Systems, Vol. 25, No. 4, pp. 46- 53, July /Aug. 2010.

[42] Zakaria Elberrichi; Karima Abidi, "Arabic Text Categorization: a Comparative Study of Different Representation Modes", International Arab Journal of Information Technology, Vol. 9, no. 5, pp. 465 – 470, Sep. 2012.