

Arabic Sentiment Analysis: A Survey

Adel Assiri^{1,2}, Ahmed Emam Ph.D^{1,3}, Hmood Aldossari¹

¹Information Systems Department, King Saud University, Riyadh, Saudi Arabia

²Abha Technology College, TVTC

³Menoufia University, Menoufia, Egypt

Abstract—Most social media commentary in the Arabic language space is made using unstructured non-grammatical slang Arabic language, presenting complex challenges for sentiment analysis and opinion extraction of online commentary and micro blogging data in this important domain. This paper provides a comprehensive analysis of the important research works in the field of Arabic sentiment analysis. An in-depth qualitative analysis of the various features of the research works is carried out and a summary of objective findings is presented. We used smoothness analysis to evaluate the percentage error in the performance scores reported in the studies from their linearly-projected values (smoothness) which is an estimate of the influence of the different approaches used by the authors on the performance scores obtained. To solve a bounding issue with the data as it was reported, we modified existing logarithmic smoothing technique and applied it to pre-process the performance scores before the analysis. Our results from the analysis have been reported and interpreted for the various performance parameters: accuracy, precision, recall and F-score.

Keywords—Arabic Sentiment Analysis; Qualitative Analysis; Quantitative Analysis; Smoothness Analysis

I. INTRODUCTION

Sentiment analysis is a type of natural language processing (NLP), where NLP or computational linguistics, is the scientific study of human languages from a computational perspective [1]. Natural language processing is an extensive field covering such applications and investigations as human language translation/generation/comprehension, speech & named entity recognition, question answering and information retrieval, word/topic segmentation, and relationship extraction. Sentiment Analysis (SA) is using natural language processing, statistics, or machine learning methods to extract, identify, or otherwise characterize the sentiment content of a text unit [2]. Sentiment analysis has also been referred to as opinion mining (OM) and is concerned with the analysis of human opinion, sentiment, and emotion about specific entities (such as food, products, organizations, etc.) and issues (politics, news, etc.) [3][4][5].

Sentiment analysis, involves in building a system to collect and examine opinions about the product made in blog posts, comments, reviews or tweets. Sentiment analysis can be useful in several ways. For example, in marketing it helps in judging the success of an ad campaign or new product launch, determine which versions of a product or service are popular and even identify which demographics like or dislike particular features [38][4][5]. This paper reviews efforts to build SA systems for Arabic. The rest of this paper has arranged as following: After a brief discussion of the

properties of Arabic language in Section 2, we review sentiment analysis process in Section 3. Related work and qualitative analysis for Arabic presented in Section 4, we presented quantitative analysis in Section 5, conclusion and future work in Section 6.

II. ARABIC LANGUAGE CHALLENGES

As an important player in international politics and the global economy, the Arab world is the focus of many multi-national interest groups and analysts who endeavour daily to decipher sentiments on issues like oil and gas prices, stock market movements, politics and foreign policy, emanating from this part of the world. The resulting chatter being in the Arabic language, there is a great need for natural language analysis of large amounts of Arabic language text and documents to support the required sentiment extraction. As described in the foregoing, the relative importance of the Arabic language in global communications demands a proportional amount of interest and research for natural-language processing of large amounts of Arabic language text and documents to facilitate sentiment extraction for industrial use [6][7][8].

The reality, however, is that there is relatively little available support for Arabic-language sentiment analysis, majorly for the following reasons: (1) relatively limited scholarly work and research funding in this area, when compared to other-language studies, especially English. (2) Morphological complexities and dialectal varieties of the Arabic language which require advanced pre-processing and lexicon-building steps beyond what is applicable for the English language domain [7][8]. This limits the potential applications of current tools and custom tools for Arabic SA may not be easy to come by, may be limited in current functionality, or may not be freely available. Farra et al [9] illustrated the challenges of Arabic-language sentiment analysis: the existence of many inflectional and derivation forms - where words have transitional meanings depending on position within a sentence, and the type of sentence (verbal or nominal). Multiple word prefixing, suffixing, affixing, and diacritical forms add high-order dimensionality for words, where the same three-letter root can generate different words in each case [9]. The nature of the Arabic language identifies the need for custom tools for Arabic SA that will be capable of identifying these diacritics and performing efficient automated POS tagging for Arabic text. As explained, morphological analyzers should be used in tandem with POS taggers to carry out root extraction as well as prefix, suffix and affix extraction. Currently, tools like MADA (Morphological Analysis and Disambiguation for Arabic) and BAMA

(Buckwalter Arabic Morphological Analyzer) are being used by Arabic language sentiment analysis researchers but these tools are far from being advanced, and there is still a need for complex and more capable POS taggers to be developed for this domain, among other issues.

III. SENTIMENT ANALYSIS PROCESS

Sentiment Analysis generally consists of three main steps: pre-processing, feature selection and sentiment classification.

A. Preprocessing

The text documents contain rich textual information such as words and phrases, punctuation, abbreviation, emoticons etc. They also tend to have misspelling, duplicate-characters (such as "coool"), especially for social media text. Direct application of SA methods on such text usually leads to poor performance. Therefore, pre-processing is typically conducted to convert the text into textual features that could be fit into the SA methods. Once the pre-processed text features are extracted, they are ready to be fit in the next phase of SA – Feature Selection [10][11]. Pre-processing is usually based on NLP techniques such as tokenization (splitting the sentences into words), de-noising (remove special characters, capture symbols for emotions), normalization (remove duplicate characters, identify root words etc.), stop-words removal (remove the stop words and the words which are of no use to sentiment analysis), stemming (return the word to its stem or root), lemmatization (convert inflected words to their root form) etc.

Haddi et al. [10] studied the role of text pre-processing in sentiment analysis, including online text cleaning, white space removal, expanding abbreviation, stemming, negation and stop words removal. For stop words, they constructed list of domain specific stop words which are not standard stop words but carry no information for the specific domain. Bao et al. [11] evaluated the effects of text pre-processing in twitter sentiment analysis. They first considered username, hashtags, emotions, digital symbols, single letters, punctuations and other non-alphabetic symbols for de-noising. Then they conducted five steps for pre-processing: URLs features reservation, negation transformation, repeated letters normalization, stemming and lemmatization. They showed that sentiment classification accuracy rises when URLs features reservation, negation transformation and repeated letters normalization are employed while descends when stemming and lemmatization are applied.

B. Feature Selection

The outputs of pre-processing are the extracted text features. Many text features are considered for SA: unigram (individual words), bigram (two consecutive words), or n-grams (n consecutive words) and either their presence for binary weighting or their frequency to indicate their relative importance; words and phrases commonly used to express opinions words and phrases commonly used to express intensification of opinions negative words that change the opinion orientation; part-of-speech (POS) to find adjectives that contains opinion information, emoticon (special characters to represent emotions). Many words in the text do not have an impact on the general orientation of it. Therefore,

keeping those words makes the dimensionality of the classification problem high and hence the classification more difficult. These words may also contain noise for the classification problem [12][13][14]. The goal of feature selection is to select important text features out of the pool of all extracted ones. Generally speaking, feature selection methods can be categorized into filter methods and wrapper methods. Filter methods rank the features according to certain metric and select the top-ranked features. Wrapper methods, on the contrary, select the best subset of features by generation and evaluation of different subsets with a classifier. Therefore, the selected features tend to be classifier specific, namely they might perform well using the specific classifier that is used for the selection, but not necessarily well with other classifiers.

The work by Yu and Wu [12] presented a 'contextual entropy model' based on basic point-wise mutual information (PMI) to perform seed word expansion originating from a small corpus of stock market news articles. The model estimates the similarity between words and seed words by comparing their relative contextual distributions using an entropy system and selecting high-match entries. Elawady et al. [13] evaluated the performance of mRMR (minimum redundancy maximum relevance), IG (information gain) and hybrid method based on Rough set theory and IG. They showed that mRMR has better performance compared with IG and the hybrid method has the best performance for sentiment analysis tasks. Agarwal and Mittal [14] considered using text features such as unigram, bigrams, the concatenation of them and POS (parts of speech). They also compared the performance of mRMR and IG and showed that mRMR is superior to IG for sentiment analysis tasks.

C. Sentiment Classification

Sentiment classification techniques are usually divided into supervised, unsupervised and semi-supervised approaches. Supervised learning uses training data to process extracted text features by adopting machine learning techniques. Unsupervised learning in the sentiment analysis context relies on robust sentiment lexicons with a sizeable number of terms with known polarity and the application of statistical-semantic weighing and distribution schemes to apply polarities to unknown words and determine the polarity of blocks of text. We can further divide unsupervised methods into dictionary-based and corpus-based relative to how the lexicon is built. [15][16][17]. Dictionary-based approach carries out a forked distributed search (two forks: antonym and synonym) for each opinion word in the dictionary. The corpus-based approach guarantees context specificity of word orientations by searching a large corpus. Lexicon-based approaches require manual collection of the opinion words and has been criticized for requiring too much human effort [15][16][17]. As a solution, the semi-supervised approach uses an initial list of seed words with annotated polarities and uses synonym-based label propagation to map polarities to unknown words [15][16][18][9].

IV. RELATED WORK

Many studies have presented several different approaches for sentiment analysis. In general, many of these studies focus on sentiment analysis for the English language and other

languages (Chinese, Italy, Ordo). There are comparatively few studies for sentiment analysis for the Arabic language. In this section we first present some important sentiment analysis studies in different languages before going on to survey the Arabic sentiment analysis studies.

A. Sentiment analysis In General

Moraes et al. [19] compared the performance of SVM (support vector machines) and NN (neural networks) for a document-level SA analysis. They showed that NN achieves better performance than SVM on balanced datasets. Rui and Liu [20] investigated pre-consumer (prior to purchase) and post-consumer (after purchase) opinion differences using NB and SVM classifiers on twitter data from both classes of users. Li and Li [21] addressed subjectivity and expresser credibility in opinion studies using SVM as the classifier. Wang et al [22] studied the performance of three popular ensemble methods (bagging, boosting, random subspace) based on five basic learners (Naive Bayes, Maximum Entropy, Decision Tree, K-Nearest Neighbour, and Support Vector Machines) on sentiment classification tasks. They showed that random subspace achieves the best results.

New developments in supervised learning show a heavy dependence on conceptual analysis. Formal Concept Analysis and Fuzzy Formal Concept Analysis (FCA/FFCA) specifically were employed in works by Li and Tsai [23] showing an abstract conceptual classification system of documents and use of training (FFCA-based conceptual classifier training as opposed to document-based training) examples to boost accuracy. Kontopoulos et al. [24] have used FCA also to build an ontology domain model. In their work, they proposed the use of ontology-based techniques toward a more efficient sentiment analysis of twitter posts by breaking down each tweet into a set of aspects relevant to the subject. Poria et al. [25] proposed a novel paradigm to concept-level sentiment analysis that merges linguistics, common-sense computing, and machine learning for improving the accuracy of tasks such as polarity detection. Yang and Cardie [26] proposed an approach that allows structured modelling of sentiment by considering both local and global contextual information. They encode intuitive lexical and discourse knowledge as expressive constraints and integrate them into the learning of conditional random field models via posterior regularization. The paper by Tang et al. [27] shows a joint sentence-level segmentation and classification system. Latent Dirichlet Allocation (LDA) was used by Xiang and Zhou [28] in the creation of topic-specific information, before going on to divide the data into several subsets based on topic distribution. In the last wave, they presented a semi-supervised training system to further increase classification accuracy. They showed that the framework can better handle the inconsistent sentiment polarity between a phrase and the words it contains. Tang et al. [29] applied neural network to learn sentiment-specific word embedding (SSWE), which encodes sentiment information in the continuous representation of words. Unsupervised approaches also have a long history for SA. Xianghua and Guo [30] presented work in the Chinese-language domain. Their work used an unsupervised approach to automatically segment Chinese social reviews into aspects - and compute the sentiment expressed in each aspect. They

used Latent Dirichlet Allocation (LDA) for aspect discovery and employed a sliding-window context over the review text to generate local topics and the linked sentiment. In [31] by Cruz and Troyano presented a taxonomy-based approach where knowledge about how people express opinions in a given domain is catalogued. They showed that this domain-specific knowledge improves opinion mining accuracy. Huang et al. [32] considered words, symbols or phrases with emotional tendencies as input features. They studied the phenomenon of polysemy in single-character emotional word in Chinese and discussed single-character and multi-character emotional word separately. Kiritchenko et al. [33] conducted SA for short informal texts on both message-level and term-level. They generated novel high-coverage tweet-specific sentiment lexicons from tweets with sentiment word hashtags and from tweets with emoticons. Pablos et al. [34] used a set of raw texts from a specific domain (the corpus) to build a list of opinion terms for that domain using seed-list propagation based on rules that featured dependency relations and POS restrictions. In unsupervised approach a significant methods are introduced in [35][36].

Semi-supervised approaches for SA have recently attracted lots of attention. A semi-supervised approach was proposed by Tang et al [37] to evaluate different types of emotional signals in Twitter data using a correlated model. The model presents dual learning based on controlled alternating propagating and fitting processes operating on labelled and unlabeled data. Zhou et al. [38] applied a semi-supervised approach Fuzzy Deep Belief Network (FDBN) on SA. The deep architecture of FDBN consists of a set of unsupervised hidden layers and a final layer of supervised training. They did a comprehensive evaluation on the state-of-the-art semi-supervised methods for SA, including semi-supervised spectral learning(Spectral), transductive SVM(TSVM), deep belief networks(DBN), personal/impersonal views(PIV), active learning(Active), mine the easy classify the hard(MECH), active deep networks(ADN), fuzzy deep belief networks(FDBN), active FDBN(AFD). A hybrid study was performed by Ortigosa et al. [39] that combined machine learning and lexicon-based approaches with a selective logic that uses machine learning when a sufficient level of labelled data is available, and a lexicon-based system when not available. They believed their approach will not only extract sentiment but also identify significant changes in emotional signatures. As we have seen in the foregoing section, there has been a lot of advancement in sentiment analysis for the English-language domain. Many highly conceptual and experimental methods have been developed to improve the performance of basic classifiers, also more work has been done to advance the scope and applicability of supervised, unsupervised, semi-supervised, and hybrid techniques. This could be the result of an abundant level of research focus in this area, as well as favorable linkages between the research and profitable industrial applications.

B. Arabic Sentiment Analysis

There are many studies have been done in opinion mining field. Most of these studies have been done in English language context, and a little in Arabic language context. In this paper we will present some studies of Arabic language

context. We present a comprehensive review of recent Arabic sentiment analysis research using a component-by-component approach.

We study the following components: approach used, methods (classifiers) used, data sources used, Arabic dialects processed, and sentiment analysis level. We also provide a merit-based assessment of the advantages and disadvantages of the sentiment analysis systems used in each research work surveyed. As we have seen in the introduction and related work, the approaches in sentiment analysis are usually divided into four classes: supervised, unsupervised, semi-supervised, and hybrid. Table 1 below categorizes the surveyed Arabic SA studies into these classes, and fig.1 shows the result.

TABLE I. ARABIC SA STUDIES BY APPROACH

S/N	Approach	Studies
1	Supervised	[1], [7], [9], [10], [11], [13], [19], [20], [27], [28], [29], [33], [41], [42], [48], [49], [52], [65], [69], [70]
2	Semi-supervised	[23], [39], [46]
3	Unsupervised	[4], [6], [21]
4	Hybrid	[5], [22], [24], [32], [51]

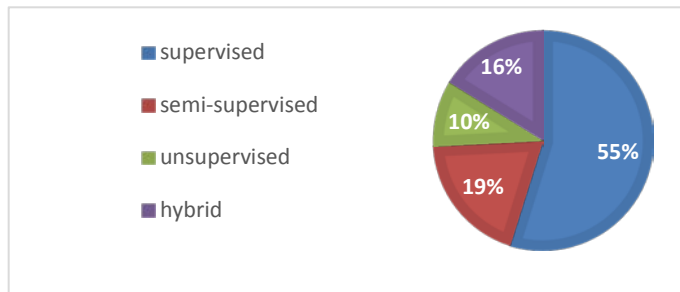


Fig. 1. Arabic SA studies by approach

From table 1, it is clear that there is a dominance of supervised learning over other techniques (semi-supervised, unsupervised, and hybrid techniques).

Arabic sentiment analysis studies used different method based on the used approach, some of these methods considered as a dominance. We have collected the various methods used in the different Arabic SA studies for supervised, semi-supervised, unsupervised and hybrid experiments and presented the result in Table 2.

From the above table, the most widely used methods (by far) appear to be based on Support Vector Machines (SVM), Naive Bayes (NB), and K-Nearest Neighbors (KNN). Lexicon-based approaches are generally prevalent across the majority of works sampled. Ensemble methods (comprising a variety of techniques) are also gaining significance.

In Arabic sentiment analysis studies several different text sources have been used, based on the objective of study, as outlined in Table 3 where the researchers in this domain appear to use Tweets, reviews/opinions & comments almost exclusively as datasets for their work on sentiment analysis. This may indicate a focus on social media.

We also investigated the size and diversity of the datasets used for the various Arabic sentiment analysis studies. We found that, there is significant variety in the quantity (size) of the datasets used in the various studies. It is more common to find studies where a single type of data was used, but there are a number of cases where multiple data types were combined.

TABLE II. ARABIC SA STUDIES BY METHOD

S/N	Method	Studies
1	SVM	[13], [24], [29], [70]
2	SVM + NB	[7], [9], [10], [11], [27], [39], [41], [42], [48], [49], [52]
3	SVM + NB + KNN	[5], [11], [19], [20], [28], [33], [69]
4	Binary Classifier	[1]
5	Maximum Entropy	[9], [22]
6	Decision Tree	[5], [9], [11], [27], [33]
7	Ensemble	[9], [11],[42]
8	Bayes Net	[9]
9	Bayes Point Machine	[29]
10	Lexicon-based	[4], [5], [6], [21], [22], [23], [24], [32], [46], [51]
11	Corpus-based	[5]
12	Grammar-based	[24]
13	Rocchio Classifiers	[42]
14	KNN (without SVM, NB)	[22]
15	NB (without SVM, KNN, D-tree)	[32], [65]

TABLE III. ARABIC SA STUDIES BY DATA

S/N	Source	Studies
1	Tweets/Twitter	[1], [5], [6], [9], [20], [21], [46], [51], [52], [65], [69], [70]
2	Wiki Pages	[1], [3]
3	Web Forums	[1], [3]
4	Reviews/Opinions	[7], [11], [19], [23], [24], [33], [39], [42], [48], [49]
5	Social Comments/Social News	[2], [4], [6], [10], [27], [65], [69], [70]
6	Lemmas	[13]
7	Website Comments	[28],[33],[41]
8	Biographic Information	[29]
9	Posts	[22], [32]
10	Documents	[24]

The Arabic language has many dialects, and no study of Arabic sentiment analysis is complete without a review of the different dialects covered in the studies. Table 4 below presents an overview of the Arabic dialect distribution in the studies surveyed. As we can see from the above table, modern Standard Arabic (MSA) sources are widely used throughout the studies sampled in this survey. Where dialects are used, Egyptian (MSA/Egyptian) was more favorable. There are also works with Levantine, Khaliji, Arabizi, Mesopotamian, Syro-Palestinian, Middle East Region, and Informal (Lebanese, Syrian, Iraqi, Libyan, Algerian, Tunisian, and Sudanese) dialects.

Also we can see that the Saudi dialect was not given attention by the researchers.

TABLE IV. ARABIC SA STUDIES BY LANGUAGE

S/N	Language	Studies
1	MSA	[1], [2], [3], [4], [5], [6], [9], [13], [19], [20], [22], [23], [24], [33], [39], [41], [42], [46], [48], [49], [65], [70]
2	MSA (Egyptian)	[7], [10], [11], [21], [27], [28], [51], [70]
3	MSA (Levantine)	[7]
4	MSA (Khaliji)	[7], [11], [65]
5	MSA (Arabizi)	[7]
6	MSA (Mesopotamian)	[11]
7	MSA (Syro-Palestinian)	[11]
8	MSA (Islamic)	[29]
9	MSA (Middle East Region)	[52]
10	MSA (Informal)	[32]

We were also interested in investigating the scope of the sentiment analysis carried out in the various Arabic language studies, so as to classify them as sentence-level, document-level, or sentence-level & document-level. Our findings are presented in Table 5 below.

TABLE V. ARABIC SA STUDIES BY PROCESSING LEVEL

S/N	Processing Level	Studies
1	Sentence-level	[2], [3], [23], [52], [69], [70]
2	Document-level	[1], [4], [5], [6], [7], [9], [10], [11], [13], [19], [20], [21], [22], [27], [28], [29], [32], [33], [39], [41], [42], [46], [48], [49], [51], [65],
3	Document-level + Sentence-level	[24]

From the previous table, it is rare to find projects in this domain that feature a combination of document-level and sentence-level sentiment analysis. On the contrary, nearly all the works sampled are focused on document-level sentiment analysis. There are also a few cases of sentence-level sentiment analysis.

One of the most crucial aspects of this work is the critical review of the various Arabic language sentiment analysis

studies surveyed, with the goal of identifying positive highlights, shortcomings, and areas of improvement, after a comprehensive review of each of the studies. Our comments are provided in Tables 6 and 7. As summarized from the above table, we have made some conclusive observations about studies in the field of Arabic Sentiment Analysis through our review of current Arabic SA research works. We found that most Arabic sentiment analysis works focus on the use of supervised methods as opposed to other classes of sentiment analysis including unsupervised, semi-supervised and hybrid or experimental systems. This method requires a huge amount of corpus and manually labeling for training and testing purpose this can be expensive, time-consuming, and difficult due to sarcasm especially in Arabic text [40][41]. The main disadvantage of this approach, it is a domain-biased which mean it give low accuracy when it is applied in different domain that was trained. This approach usually use machine learning methods such as Support Vector Machines, Naïve Bayes Classifiers and Maximum Entropy approaches [40][41]. In the other hand some studies employed the lexicon-based approach using different techniques to generate sentiment lexicons that would contribute to the task of sentiment analysis. This approach is based on a list of sentiment words with their polarities to determine the sentiment of review. This approach is considered practical since it is not domain-biased, recently some researchers intended to use the ontology in this approach, and such ontology may be used for different tasks: Arabic NLP tools, information retrieval [42]. Dialects are not supported in many of the Arabic SA studies surveyed in this paper. This presents a major disadvantage because the Arab language is dialectically rich and its diverse structural properties in the various dialects need to be fully captured in order to derive maximum benefit from Arabic SA, especially for less-formal channels like Social Media, whose corpora are principally not in Modern Standard Arabic (MSA).

It was also noticed that a limited set of classifiers (techniques) were repeatedly used for sentiment analysis in many of the papers surveyed. While researchers probably choose these same set of classifiers because they are proven to be effective, value is not being added to the field of Arabic SA if more experimental or conceptually novel techniques are not implemented or investigated. There is very little focus on sentence-level sentiment analysis for many of the studies. Most of the observations recorded during this survey generally lead to the conclusion that Arabic sentiment analysis is in its growing phases.

TABLE VI. ADVANTAGES ANA DISADVANTAGES SUMMARY

Paper ID	Advantages	Disadvantages
[1]	showed extensive list of features, studied the importance of different features	disregarded neutral and mixed classes
[2]	The annotations are extensive	No Sentiment Analysis evaluations on the corpus
[3]	Multi-genre corpus	No Sentiment Analysis evaluations on the corpus
[4]	Multiple lexicons constructed, integrated lexicon achieves best performance	Dialects are not considered
[5]	Negation and Intensification are considered	Neutral class is not included. Sarcasm is not considered
[6]	Advanced lexicon construction	Using individual words polarities technique
[7]	showed extensive list of features, studied the importance of different features	could try more classification methods, no details given on how sampling is conducted to obtain a balanced subset of data
[9]	Pre-processing leads to improvement	Tags need to be added manually
[10]	Introduced Social Network specific features	Dialects are not considered
[11]	Besides classification on subjectivity and polarity, also considered intensity classification	Does not deal with Emoticons, chat language and Arabizi
[13]	Large-scale lexicon	Could try more classifiers, dialects not considered
[19]	Studied the effects of pre-processing and the characteristics of the dataset	Could try more classifiers, dialects not considered
[20]	Developed three lexicons as well as a negation library, the dataset was large	Intensifications were not considered
[21]	Evaluated methods to learn the weights of the words and combine such weights	Dialects are not considered
[22]	Combination of multiple methods improves the performance	Considered posts from only three domains
[23]	label propagation is effective for lexicon construction	Only considered sentence level
[24]	Considered both grammar and lexicon	Dialects, suffix and prefix extraction not extracted, small dataset
[27]	Developed three lexicons as well as a negation library, the dataset was large	Only sentence level
[28]	Particularly addressed slang language	No benefits for non-slang cases
[29]	NLP is used, word presence feature leads to better performance	Could use more classifiers

TABLE VII. ADVANTAGES AND DISADVANTAGES SUMMARY

Paper ID	Advantages	Disadvantages
[32]	Considered both supervised and unsupervised approaches	Evaluated limited supervised methods
[33]	SentiStrength has better performance than SocialMention	Dialects are not considered
[39]	Semi-supervised lexicon construction	Dialects, Franco Arabic and compound phrases are not considered (single word match only)
[41]	Addressed unbalanced classification	Proposed methods didn't show advantage
[42]	Ensemble classifier achieves better classification	More classifiers can be added to the ensemble
[46]	Extensive feature categories, addressed topic shift	semi-supervised approach improves subjectivity analysis but not sentiment analysis
[48]	The Corpus has good quality	Could include more features
[49]	Determines the polarity of an Arabic corpus using English translation	SA depends on the quality of the translation
[51]	Very detailed investigation on the processing techniques	pre-processing techniques could be improved by cross-validation, lexicon might not be extensive
[52]	n -gram features are used	The corpus is small and low frequency terms are ignored
[65]	Sizeable dataset used	Could have used more classifiers
[69]	Used an ensemble of classifiers with a relatively comprehensive dataset	The size of the dataset is small
[70]	Supported dialectal Arabic in addition to MSA	Could have used more classifiers, relatively limited dataset

V. QUANTITATIVE ANALYSIS OF RECENT ARABIC SA RESEARCH

Our primary concern for performing a quantitative analysis on the performance data provided by the different Arabic sentiment analysis studies is to determine if, and the degree to

which, there is any significant difference in the performance outputs (evaluated across accuracy, precision, recall and F-score) for each of the methods used in the research works being surveyed, as this knowledge will put us in a position to potentially identify areas for improvement in current approaches. Table 8, catalogues reported statistics collated

from the various research publications being surveyed. Note: where multiple results were provided in these works, we selected only the best results. Every attempt was made to state the results as they were originally published by their various authors.

TABLE VIII. REPORTED STATISTICS FROM SURVEYED STUDIES

Paper ID	Accuracy	Precision	Recall	F-score
[1]	95.83%			
[4]	74.60%			
[6]	61.20%	60.60%	63.90%	
[9]	87.43%			
[10]	61.40%			
[11]	96.90%	95.00%	97.00%	
[13]				71.10%
[19]	97.20%	99.60%	94.80%	
[20]	76.78%			
[21]	83.80%	44.40%	57.10%	49.00%
[22]	84.34%	87.20%	89.62%	85.57%
[24]	87.00%			
[27]		87.40%	33.80%	
[28]		88.60%	78.00%	88.54%
[29]	95.91%			
[32]	91.20%			
[33]	99.20%			
[39]	97.81%	98.00%	98.00%	
[41]	96.00%	98.00%	98.00%	
[42]		98.60%	98.60%	97.60%
[49]		90.00%	95.00%	90.73%
[51]	75.90%	76.90%	75.90%	76.20%
[65]	80.60%	86.10%	99.90%	83.20%
[69]			83.00%	72.00%
[70]	99.90%	99.90%	99.90%	99.90%

A. Analysis Technique – Smoothness Analysis

Smoothness analysis is based on arithmetic series in discrete mathematics [42]. For any arithmetic series, we have a first term S_1 , last term S_n and common difference d such that any member of the series can be represented as:

$$S_i = S_1 + (i - 1)d$$

Because real-life data may not always behave as an arithmetic series, the smoothness of a distribution is simply an estimation of the error in the real distribution relative to the projected arithmetic series distribution [42]:

$$\text{smoothness} = \left(1 - \frac{S_1 + S_n}{2\bar{S}}\right) \quad (1)$$

Where: S_1 = first term in the real data series when arranged in increasing order, S_n = last term in the real data series when arranged in increasing order, and \bar{S} = average of the real data series.

Benefit: the smoothness of a distribution as calculated by equation (1) gives us the % error (percentage error) in the straight-line form of the data, and tells us how the data has

changed with respect to the different input values (that is, we can evaluate the impact or significance of the different methods used by the research works on the performance scores reported).

B. Local Optimum Problem in Smoothness Analysis

When evaluating the impact of studies using equation (1) and the data from Table 7, we run into the problem of local optimum: an approximately constant score for all performance categories. This is because all the performance scores are less than 1 and are therefore, similar.

Lemma: For all pairs of similar values, the smoothness function will return a zero (no impact) result.

$$\forall (S_1, S_2) : S_1 \approx S_2 \approx S, \lim \left(1 - \frac{S_1 + S_2}{2\bar{S}}\right) \rightarrow \left(1 - \frac{2S}{2\bar{S}}\right) \rightarrow 0$$

This invalidates the analysis unless a solution can be obtained to proportionally amplify the input values (performance scores), so that the validity condition (shown below) can be met:

$$\lim \left(1 - \frac{S_1 + S_2}{2\bar{S}}\right) \neq 0$$

(Validity condition for smoothness function)

C. Solution to Local Optimum: Logarithmic Smoothing

To solve the local optimum problem described above which will invalidate our analysis according to Lemma due to the closely-bounded performance scores, we explore the use of the logarithmic smoothing technique described in [43], a procedure for proportionally expanding individual elements within the space of a closely-bounded range.

$$\Gamma(\ln r, \ln \theta, \ln \phi) = r' \theta' \phi' e^{\frac{r}{r_{\max}} + \frac{\theta}{\theta_{\max}} + \frac{\phi}{\phi_{\max}} - 3} \quad (2)$$

The logarithmic smoothing process is shown above equation (2) for a three-dimensional smoothing problem in spherical coordinates (r, θ, ϕ).

Where: r = component of a point in r -coordinate, θ = component of a point in θ -coordinate, ϕ = component of a point in ϕ -coordinate, r' = target projection of r , θ' = target projection of θ , ϕ' = target projection of ϕ , r_{\max} = maximum value of r , θ_{\max} = maximum value of θ , ϕ_{\max} = maximum value of ϕ

Benefit: At any point within the sphere (3D space), the function $\Gamma(\ln r, \ln \theta, \ln \phi)$ gives a smooth projection that is continuous in r, θ , and ϕ directions [43].

This means that with this transformation, the problem of local optimum can be reasonably avoided because input values are transformed to their smooth projections $r \rightarrow r_{\text{smooth}}, \theta \rightarrow \theta_{\text{smooth}}, \phi \rightarrow \phi_{\text{smooth}}$ and these values will pass the validity condition because $r_{\text{smooth}} > r, \theta_{\text{smooth}} > \theta, \phi_{\text{smooth}} > \phi$ and $\frac{r_{\text{smooth}}}{r} \neq \frac{\theta_{\text{smooth}}}{\theta} \neq \frac{\phi_{\text{smooth}}}{\phi}$ such that we have a valid analysis (by the validity condition for smoothness function).

For our purpose in this analysis, we present a simplification of this idea as follows:

As we only have 1-dimensional data (each performance parameter is evaluated on a case-by-case basis – accuracy only, precision only, recall only and F-score only), for which only the r-coordinate is sufficient, we need to remove unnecessary coordinates (θ, ϕ) by setting these values to 1: $\theta = \theta_{\max} = 1, \phi = \phi_{\max} = 1$

This reduces equation (2) to a form that is applicable for our analysis, which is:

$$\Gamma(\ln r, \ln 1, \ln 1) = \Gamma(\ln r, 0, 0) = r'(1)(1)e^{\frac{r}{r_{\max}}+1+1-3} = r'e^{\frac{r}{r_{\max}}-1}$$

Which we can write as:

$$\Gamma(\ln r) = r'e^{\frac{r}{r_{\max}}-1} \quad (3)$$

Conclusion: equation (3) above is the logarithmic smoothing function that we will use in our analysis to solve the problem of local optimum.

Fig.2 below shows the effect of applying the logarithmic smoothing function equation (3) in transforming data from closely-bounded spaces (x-space) to loose-bounded spaces (L-space):

Outcome: there is proportional amplification in the data, such that the behavior of the data remains unchanged, while small differences are much easily visualized and evaluated.

D. Results of The Analysis

Comparative Results – Accuracy: Table 9 shows the raw accuracy scores and the converted logical scores for use in the smoothness analysis.

To arrive at the logical scores shown in the table above (used for the analysis), we used the function of logarithmic smoothing, by setting $r' = 1000$, calculating converted scores and arranging in increasing order. For this table, max: r (largest element in r) is 0.999.

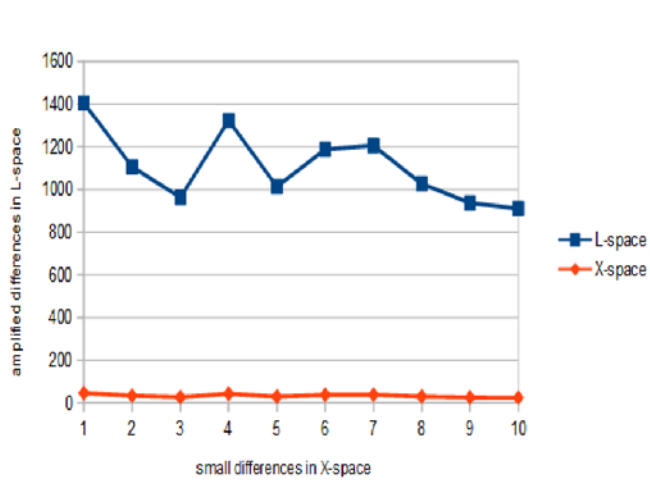


Fig. 2. Transformation from closely-bound x-space to widely-bound l-space

The smoothness = 0.0463, for this dataset (*slightly rough*), as calculated by using the smoothness function, indicating that: there is no significant impact of the different methods used on the accuracy. See Fig. 3 for a visualization (plot correlates well with trend-line). By using the same process, we obtained: 0.10973023, 0.12700562, and 0.045964546, as smoothness result for precision, recall, and F-score, respectively. The results lead to the following conclusions: there is slight impact of the different methods used on the precision and recall (see *slightly rough* curves in Fig.3, Fig.4, and Fig.5 – the plots do not correlate very well with their trend-lines) but there is no significant impact on accuracy and F-score (see *smooth* curves in Fig. 3, and Fig.4 – the plots correlate well with their trend-lines).

Table 10 presents a summary of smoothness results obtained from the experiments. As can be seen from this analysis, *Accuracy & F-score* are not impacted by the different methods adopted by the various researchers in the studies surveyed. *Precision & Recall*, however, show slight response to the different methods used by the researchers in the studies surveyed. **Table 8:** Accuracy distribution by logical score and raw values

TABLE IX. ACCURACY DISTRIBUTION BY LOGICAL SCORE AND RAW VALUES

Study	raw accuracy score	Converted logical score
[6]	0.6120	678.828
[10]	0.6140	680.188
[4]	0.7460	776.271
[51]	0.7590	786.438
[20]	0.7678	793.397
[65]	0.8060	824.322
[21]	0.8380	851.154
[22]	0.8434	855.768
[24]	0.8700	878.860
[9]	0.8743	882.651
[32]	0.9120	916.597
[1]	0.9583	960.078
[29]	0.9591	960.847
[41]	0.9600	961.713
[11]	0.9690	970.416
[19]	0.9720	973.334
[39]	0.9781	979.296
[33]	0.9920	993.017
[70]	0.9990	1000

TABLE X. SUMMARY OF SMOOTHNESS RESULTS

Slight significance	No significance
Precision (0.1097)	Accuracy (0.0463)
Recall (0.1270)	F-score (0.0459)

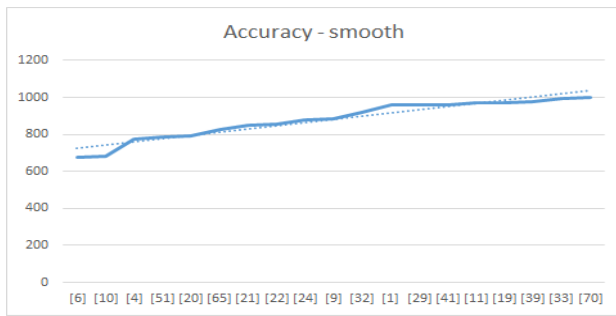


Fig. 3. Smoothness accuracy result

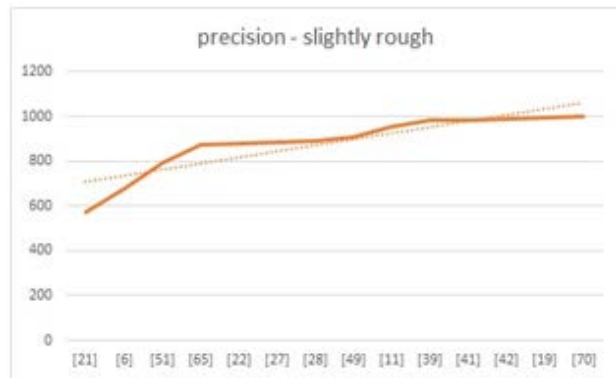


Fig. 4. Smoothness precision result

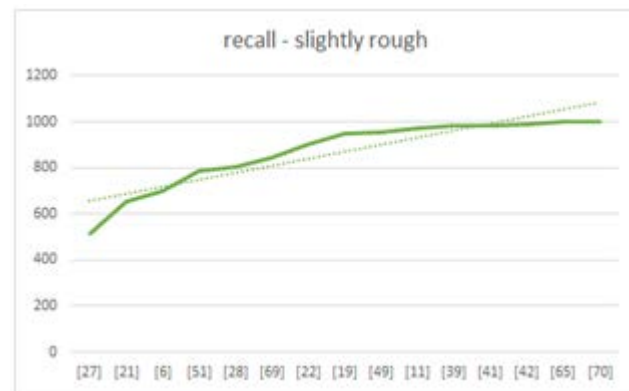


Fig. 5. Smoothness recall result

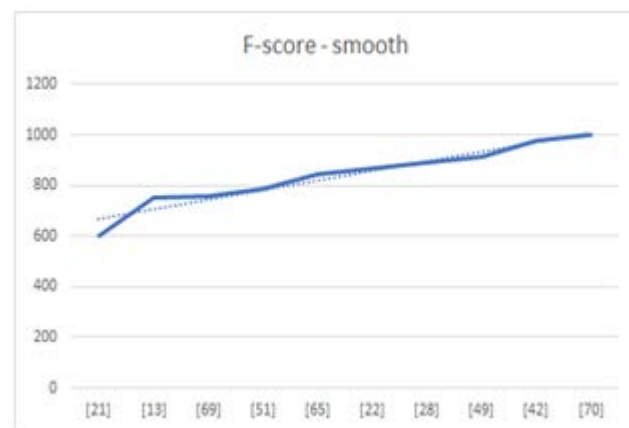


Fig. 6. Smoothness f-score result

VI. CONCLUSION AND FUTURE WORK

In this paper we have surveyed the important Arabic sentiment analysis studies qualitatively and quantitatively. We have presented detailed analyses of methods used and results obtained in the current Arabic sentiment analysis studies, as well as a rich discourse on the direction of current research, present limitations. In our qualitative evaluation, we found that, the majority of Arabic SA uses established supervised methods as opposed to more progressive or experimental unsupervised and semi-supervised approaches. The dialects are not processed in many of the Arabic SA studies surveyed, which is a major drawback on the effectiveness of current Arabic SA because most of the available Arabic language text in the social media and other spaces represent a wide range of distinct, autonomous, and morphologically complex Arabic language dialects. It was also observed that many of the studies surveyed used the same limited set of classifiers - raising questions about reasonable value added to the field if every study essentially repeats the same experiment on a Different dataset. There is a definite need for more inventiveness and creativity in the design of experiments as well as the development of novel classification and analysis techniques beyond the established algorithms.

For our quantitative evaluation, we applied rigorous data modelling and statistical procedures to investigate the effectiveness of methods adopted by the various researchers in the Arabic SA works surveyed. We collected performance data (accuracy, precision, recall and f-score) for the various studies and applied advanced techniques including logarithmic smoothing field analysis and a relative smoothness function, to uncover deep patterns in the performance data. Our approach was based on the reasoning that similar processes will produce similar results. The various studies conducted for Arabic SA will not be differentiable if they all produce similar results across the various performance classes - accuracy, precision, recall and f-score. But where we have significant variance of results, then there is opportunity for improvement. The analysis performed yielded the following conclusion: there is only a slight impact of the different methods used on the Precision & Recall of results obtained while there was no significant impact on the Accuracy & F-score. This ultimately leads us to the conclusion that Arabic SA researchers should employ a more diverse set of techniques and approaches that do more to improve scoring across the full range of performance parameters.

In the future work, we believe that there is a promising trend to obtain optimal Arabic SA system. We intend to propose and develop a new hybrid method using deep learning technique and big data technique such as Hadoop and MapReduce to solve some of the existing problems in Arabic sentiment analysis as highlighted in this survey as well as to obtain optimal system for Arabic SA. As we have seen, most of the work in the field of Arabic sentiment analysis has focused on the use of supervised learning techniques, and are largely lexicon-based approaches with the characteristic limitations. We believe that the opportunity space for growth in this field will be driven by the exploration of unsupervised learning techniques, principally through hybrid method.

REFERENCES

- [1] Association for computational linguistics. What is Computational Linguistics? <http://www.aclweb.org/archive/misc/what.html> (accessed 18 September 2015).
- [2] Web. Introduction to Sentiment Analysis, <http://lct-master.org/files/mullensentimentcourseslides.pdf> (accessed 18 September)
- [3] Liu B. Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*; 5(1): pp. 1-167, 2012.
- [4] Tsytarou M and Palpanas T. Survey on mining subjective data on the web. *Data Mining and Knowledge Discovery*; 24(3): pp. 478-514, 2012.
- [5] Spanish society for natural language processing. Natural Language Processing and Opinion Mining, http://www.aepia.org/aepia/files/evia/transparencias/urea_evia2014.pdf (accessed 18 September 2015).
- [6] Al-kabi MN, Abdulla NA and Al-ayyoub M. An analytical study of arabic sentiments: maktoob case study. In: 8th international conference for internet technology and secured transactions, IEEE, London, UK, , pp. 89-94, 2013.
- [7] Sarah O. Alhumoud, Mawaheb I. Altuwajri, Tarfa M. Albuhairei, Wejdan M. Alohaideb. Survey on Arabic Sentiment Analysis in Twitter. *International Science Index*, , 9 (1), pp. 364-368, 2015.
- [8] Korayem M, Crandall D and Abdul-mageed M. Subjectivity and sentiment analysis of arabic: a survey. In *Advanced Machine Learning Technologies and Applications*, , 128-139, 2012.
- [9] Farra N, Challita E, Abou-assi R and Hajj H. Sentence-level and document-level sentiment mining for arabic texts. In: International conference on data mining workshops, IEEE, , pp. 1114-1119, 2010.
- [10] Haddi E, Liu X and Shi Y. The role of text pre-processing in sentiment analysis. *Procedia Computer Science*, pp. 26-32, 2013.
- [11] Bao Y, Quan C, Wang L and Ren F. The role of pre-processing in twitter sentiment analysis. *The 10th International Conference on Intelligent Computing*, Taiyuan, China.; pp. 615-624, 2014.
- [12] Yu LC, Wu JL, Chang PC and Chu HS. Using a contextual entropy model to expand emotion words and their intensity for the sentiment classification of stock market news. *Knowledge-based Systems*, 41, pp. 89-97, 2013.
- [13] Arafat H, Elawady RM, Barakat S and Elrashidy NM. Different feature selection for sentiment classification. *International Journal of Information Science and Intelligent Systems*; 3(1): pp. 137-150, 2014.
- [14] Agarwal B and Mittal N. Optimal feature selection for sentiment analysis. In: *Computational linguistics and intelligent text processing*, Springer, , pp. 13-24, 2013.
- [15] Abdulla N, Majdalawi R, Mohammed S, Al-ayyoub M and Al-kabi M. Automatic lexicon construction for Arabic sentiment analysis. *International Conference on Future Internet of Things and Cloud (FiCloud-2014)*, Barcelona, Spain, , pp. 547-552, 2014.
- [16] Abdulla NA, Ahmed NA, Shehab MA and Al-ayyoub M. Arabic sentiment analysis: lexicon-based and corpus-based. In: *Applied electrical engineering and computing technologies*, IEEE Jordan Conference, , pp. 1-6, 2013.
- [17] Abdulla NA, Ahmed NA, Shehab MA, Al-ayyoub M, Al-kabi MN and Al-rifai S. Towards improving the lexicon-based approach for arabic sentiment analysis. *International Journal of Information Technology and web Engineering*; 9(3): pp. 55-71, 2014.
- [18] El-beltagy SR and Ali A. Open issues in the sentiment analysis of arabic social media: a case study. In: 9th international conference on innovations in information technology, IEEE, , pp. 215-220, 2013.
- [19] Moraes R, Valiati JF and Neto WPG. Document-level sentiment classification: an empirical comparison between SVM and ANN. *Expert Systems with Applications*; 40(2): pp. 621-633, 2013.
- [20] Rui H, Liu Y and Whinston A. Whose and what chatter matters? the effect of tweets on movie sales. *Decision Support Systems*; 55(4): pp. 863-870, 2013
- [21] Li YM and Li TY. Deriving market intelligence from microblogs. *Decision Support Systems*; 55(1): pp. 206-217, 2013.
- [22] Wang G, Sun J, Ma J, Xu K and Gu J. Sentiment classification: the contribution of ensemble learning. *Decision Support Systems*; 57, pp. 77-93, 2014.
- [23] Li ST and Tsai FC. A fuzzy conceptualization model for text mining with application in opinion polarity classification. *Knowledge-based Systems*; 39, pp. 23-33, 2013.
- [24] Kontopoulos E, Berberidis C, Dergiades T and Bassiliades N. Ontology-based sentiment analysis of twitter posts. *Expert Systems with Applications*; 40(10): pp. 4065-4074, 2013.
- [25] Poria S, Cambria E, Winterstein G and Huang GB. Sentic patterns: dependency-based rules for concept-level sentiment analysis. *Knowledge-based Systems*; 69, pp. 45-63, 2014.
- [26] Yang B and Cardie C. Context-aware learning for sentence-level sentiment analysis with posterior regularization. In: *Proceedings of the ACL*, Baltimore Maryland, , pp. 325-335, 2014.
- [27] Tang D, Wei F, Qin B, Dong L, Liu T and Zhou M. A joint segmentation and classification framework for sentiment analysis. In: *Proceeding of EMNLP conference*, 2014.
- [28] Xiang B and Zhou L. Improving twitter sentiment analysis with topic-based mixture modeling and semi-supervised training. In: *Proceedings of the 52nd annual meeting of the association for computational linguistics*, ACL, , pp. 434-439, 2014.
- [29] Tang D, Wei F, Yang N, Zhou M, Liu T and Qin B. Learning sentiment-specific word embeddings for twitter sentiment classification. In: *proceedings of the 52nd annual meeting of the association for computational linguistics*, ACL, pp. 1555-1565, 2014.
- [30] Xianghua F, Guo L, Yanyan G and Zhiqiang W. Multi-aspect sentiment analysis for chinese online social reviews based on topic modeling and hownet lexicon. *Knowledge-based Systems*; 37, pp. 186-195, 2013.
- [31] Cruz FL, Troyano JA, Enriquez F, Ortega FJ and Vallejo CG. Long autonomy or long delay? the importance of domain in opinion mining. *Expert Systems and Applications*; 40(8): pp. 3174-3184, 2013.
- [32] Huang Z, Zhao Z, Liu Q and Wang Z. An unsupervised method for short-text sentiment analysis based on analysis of massive data. In: *Intelligent computation in big data era*, Springer, , pp. 169-176, 2015.
- [33] Kiritchenko S, Zhu X and Mohammad SM. Sentiment analysis of short informal texts. *Journal of Artificial Intelligence Research*; 50, pp. 723-762, 2014.
- [34] García-Pablos A., Cuadros M., Gaines, S. Rigau G. V3: Unsupervised Acquisition of Domain Aspect Terms for Aspect Based Sentiment Analysis. 8th International Workshop on Semantic Evaluation (SemEval 2014) at Third Joint Conference on Lexical and Computational Semantics and the 25th International Conference on Computational Linguistic (COLING'14). Dublin, Ireland. 2014.
- [35] Hassan S, Miriam F, Yulan H and Harith A. Senticircles for contextual and conceptual semantic sentiment analysis of twitter. In: *proceedings of the 11th ESWC conference*, Crete, Greece, 2014.
- [36] Hassan S, Yulan H and Harith A. Semantic sentiment analysis of twitter. In: *proceedings of the 11th international semantic web conference*, Boston, USA, 2012.
- [37] Tang J, Nobata C, Dong A, Chang Y and Liu H. Propagation-based sentiment analysis for micro-blogging data. In: *SIAM international conference on data mining*, 2015, Vancouver, British Columbia, Canada. pdf [185].
- [38] Zhou S, Chen Q and Wang X. Fuzzy deep belief networks for semi-supervised sentiment classification. *Neurocomputing*; 131, pp. 312-322, 2014.
- [39] Ortigosa A, Mart JM and Carro RM. Sentiment analysis in facebook and its application to e-learning. *Computers in Human Behavior*; 31, pp. 527-541, 2014.
- [40] Abdul-mageed M, Diab M and Kubler S. Samar: subjectivity and sentiment analysis for Arabic social media. *Journal of Computer Speech & Language*; 28(1): pp. 20-37, 2014.
- [41] Al-kabi MN, Alsmadi IM, Gigieh AH, Wahsheh HA and Haidar MM. Opinion mining and analysis for arabic language. *International Journal of Advanced Computer Science and Applications*; 5(5), pp. 181-195, 2014.

- [42] Maha A, Mona A, Nehal A, Wafa A, Asma A and Mesheal A. Semtree ontology for enriching arabic text with lexical semantic annotations. In: Proceedings of the 9th conference on semantic computing, IEEE, 2015.
- [43] Schmidt P and Ayres F. Schaum's outline of college mathematics. 4th edition. McGraw-Hill, 2010.
- [44] Spiegel MR. Schaum's Theory and Problems of Theoretical Mechanics. McGraw-Hill, 1967.
- [45] Abdul-mageed M and Diab MT. Subjectivity and sentiment annotation of modern standard Arabic newswire. In: Proceedings of the 5th linguistic annotation workshop, Association for Computational Linguistics, , pp. 110-118, 2011.
- [46] Abdul-mageed M and Diab MT. Awatif: a multi-genre corpus for modern standard arabic subjectivity and sentiment analysis. In: proceeding of the 8th International conference on language resources and evaluation (LREC), pp. 3907-3914, 2012.
- [47] Abdulla NA, Al-ayyoub M and Al-kabi MN. An extended analytical study of arabic sentiments. International Journal of big Data Intelligence; 1(1): pp. 103-113, 2014.
- [48] Ahmed S, Pasquier M and Qadah G. Key issues in conducting sentiment analysis on arabic social media text. In: 9th international conference on innovations in information technology, IEEE, , pp. 72-77, 2013.
- [49] 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 ANLP 2014, EMNLP 2014 Doha, Qatar, , p. 165, 2014.
- [50] Bai X. Predicting consumer sentiments from online text. Decision Support Systems; 50(4): pp. 732-742, 2011.
- [51] Deng L and Dong Y. Deep learning: methods and applications. Foundations and Trends in Signal Processing; 7(3-4): pp. 197-387, 2014.
- [52] Ding C and Peng H. Minimum redundancy feature selection from micro-array gene expression data. Journal of Bioinformatics and Computational Biology; 3(2): pp. 185-205, 2005.
- [53] Duwairi R and El-orfali M. A study of the effects of preprocessing strategies on sentiment analysis for arabic text. Journal of Information Science; 40(4): pp. 501-513, 2014.
- [54] Duwairi RM, Marji R, Sha'ban N and Rushaidat S. Sentiment analysis in arabic tweets. In: 5th international conference on information and communication systems, IEEE, , pp. 1-6, 2014.
- [55] El-halees A. Arabic opinion mining using combined classification approach. In: proceeding of the international Arab Conference on Information Technology. ACIT (2011).
- [56] Elhawary M and Elfeky M. Mining arabic business reviews. In: International conference on data mining workshops, IEEE, pp. 1108-1113, 2010.
- [57] A. E.-D. A. Hamouda and F. E. El-taher. Sentiment Analyzer for Arabic Comments System. International journal Adv. Comput. Sci. Appl., , 4 (3), pp. 100–103, 2013.
- [58] Taysir Hassan A, Soliman, M, Ali M, Abdel Rahman Hedar, M. M. Doss, "MINING SOCIAL NETWORKS' ARABIC SLANG COMMENTS". In Proceedings of IADIS European Conference on Data Mining (ECDM'13), Prague, Czech Republic, 2013.
- [59] Helmy T and Daud A. Intelligent Agent for Information Extraction from Arabic Text without Machine Translation. CEUR, Vol. 687, 2010.
- [60] Itani MM, Hamandi L, Zantout RN and Elkabani I. Classifying sentiment in arabic social networks: naive search versus naive bayes. In: 2nd international conference on advances in computational tools for engineering applications, IEEE, , pp. 192-197, 2012.
- [61] Khasawneh RT, Wahsheh HA, Al-kabi MN and Alsmadi IM. Sentiment analysis of arabic social media content: a comparative study. In: International conference on information science and technology, IEEE, 2013, pp. 101-106.
- [62] Mahyoub FHH, Siddiqui MA and Dahab MY. Building an arabic sentiment lexicon using semi-supervised learning. Journal of King Saud University-computer and Information Sciences; 26(4): pp. 417-424, 2014.
- [63] Moutassir A, Benbrahim H and Berrada I. Some methods to address the problem of unbalanced sentiment classification in an arabic context. In: Colloquium in information science and technology, IEEE, , pp. 43-48, 2012.
- [64] Omar NO, Albared M, Al-shabi A and Al-moslimi T. Ensemble of classification algorithms for subjectivity and sentiment analysis of arabic customer reviews. International Journal of Advancements in Computing Technology; 14(5): pp. 77-85, 2013.
- [65] Pearl J. Probabilistic reasoning in intelligent systems: networks of plausible inference. Morgan Kaufmann Publishers Inc. San Francisco, CA, USA ©1988
- [66] Refaee, E. and Rieser, V. An Arabic Twitter Corpus for Subjectivity and Sentiment Analysis. The 9th edition of the Language Resources and Evaluation Conference LREC'2014. The European Language Resources Association. Reykjavik, Iceland, May, pp.26-31, 2014.
- [67] Rushdi-saleh M, Martin-valdivia MT, Urena-lopez LA and Perea-ortega JM. Oca: opinion corpus for arabic. Journal of the American Society for Information Science and Technology; 62(10): pp. 2045-2054, 2011.
- [68] Rushdi-saleh M, Martin-valdivia MT, Urena-lopez LA and perea-ortega JM. Bilingual experiments with an Arabic-english corpus for opinion mining. In Proceedings of Recent Advances in Natural Language Processing, Hissar, Bulgaria, pp.740-745, 2011.
- [69] Safavian AR and Landgrebe D. A survey of decision-tree classifier methodology. IEEE Transactions on Systems, man and Cybernetics; 21(3): pp. 660-674, 1991.
- [70] Shoukry A and Rafea A. Preprocessing egyptian dialect tweets for sentiment mining. In: Proceedings of the 4th workshop on computational approaches to arabic script-based languages, p. 47, 2012.
- [71] Shoukry A and Rafea A. Sentence-level arabic sentiment analysis. In: International conference on collaboration technologies and systems, IEEE, , pp. 546-550, 2012.
- [72] Specht DF. Probabilistic neural networks. Neural Networks; 3(1): pp. 109-118, 1990.
- [73] Mourad A and Darwish K. Subjectivity and sentiment analysis of modern standard arabic and arabic microblogs. In: Proceedings of the 4th workshop on computational approaches to subjectivity, sentiment and social media analysis, , pp. 55-64, 2013.
- [74] Duwairi R and Qarqaz I. Arabic sentiment analysis using supervised classification. In: 1st international workshop on social networks analysis management and security, 2014.
- [75] Hossam I, Sherif A and Mervat G. Sentiment analysis for modern standard arabic & colloquial. International Journal on Natural Language Computing; 4(2): pp. 95-109, 2015.
- [76] Sallab A, Hajj H, Badaro G, Baly R, El-hajj W and Shaban K. Deep learning models for sentiment analysis in arabic. In: Proceedings of the 2nd workshop on arabic natural language processing, Beijing, China. Association for Computational Linguistics, pp. 9-17, 2015.