IJSBAR International Journal of Sciences: Basic and Applied Research

International Journal of Sciences: Basic and Applied Research (IJSBAR)

ISSN 2307-4531
(Print & Online)

http://gssrr.org/index.php?journal=JournalOfBasicAndApplied

International Journal of Sciences: Basic and Applied Research
ISSN 2307-4531
(Print & Online)
Published by: GSSRR

Sentiment Analysis in Arabic: An Overview

Saud S. Alotaibi*

Umm Al-Qura University, Makkah, Saudi Arabia

Email: ssotaibi@uqu.edu.sa

Abstract

The analysis of natural language text for identification of sentiment has been well-studied for the English language. In contrast, the work that has been done in Arabic remains in its infancy; thus, requiring more cooperation between research communities to offer a mature sentiment analysis system for Arabic. There are recognized challenges that face linguists in this field; some of them inherited from the nature of the Arabic language itself, and others derived from the scarcity of tools and sources. This article provides an overview of sentiment analysis in the Arabic language, by detailing what has been done in English as a model example of such an analysis, and what have been covered to date in Arabic, as well as some of the limitations and existing potential research avenues in this field.

Keywords: Natural Arabic Language Processing; Arabic Sentiment Analysis; Arabic Sentiment Classification; Subjectivity Classification; Polarity Classification.

1. Introduction

Over the past decade, the extraction of sentiment from text has attracted a lot of attention, both in industry and in academia. Sentiment analysis attempts to establish people's feeling from their writing. Many fields are included in this topic, such as natural language processing, machine learning, and computational linguistics. A lot of this research has been done in English [1], as this is the dominant language of the science. Recently, a few researchers have concentrated on applying sentiment analysis to other languages, one such language being Arabic.

---

* Corresponding author.

Arabic is not like European languages, such as English because of its richer morphological structure. It also has many challenges which require special treatment. Therefore, Arabic natural language processing has become attractive to researchers due to its complexity and the scarcity of available resources; as a result, the importance of addressing this language has been noted. These days, it can be seen that strong effort is being made with the fundamental tools of Arabic, such as the morphological analyzer, part-of-speech tagger and syntactical parser, according to A. Farghaly and K. Shaalan [2], but the field is still at an early stage of evolution. Nevertheless, an upper level of applications, such as Sentiment Analysis (SA), is still in the beginning stages, and more effort and reliable low-level tools are required in order to build on the top of this foundation.

The following sections are organized as follows. The next section shows a general description and assessment of the sentiment analysis field. The methodologies including techniques, features, and corpora of sentiment analysis will be discussed in section 2. Section 3 will provide some background regarding the Arabic language as well as the challenges in Arabic natural processing. Section 4 illustrates the related works of Sentiment Analysis in Arabic language then followed by the discussion in section 5. Some possible research directions are given in section 6. This article ends with a conclusion.

2. Sentiment Analysis

Sentiment Analysis SA is a method of capturing the sentiment (feeling or opinion) of people towards a specific topic. This field may be considered part of the following areas: machine learning, natural language processing and computational linguistics. In other words, it usually tries to evaluate and extract the sentiment of people from their writing. In literature, SA has many names, including subjectivity analysis, opinion mining, review mining and appraisal extraction (B. Pang and L. Lee 2008). Moreover, the sentiment of a text can be explicit or implicit. If explicit, a text directly gives a sentiment, such as (إنها سيارة رائعة /InhA syArħ rAŷçħ/ “It is a nice car”), it should be noted that throughout this report, Arabic words are represented in three variants: (Arabic word / transliteration scheme [3] / “English translation”), while if implicit the text implies a kind of sentiment, like (عمل الشاحن لمدة اسبوع فقط /çml AlšAHn lmdħ Asbwç fqT/ “The charger only worked for one week”). More formally, SA can be defined as:

Given a text t from a text set T, computationally assigning polarity labels p from a set of polarities P in such a way that p would reflect the actual polarity that is found in t.

In the subjectivity analysis, the first aim is to determine or classify whether the the content of the text is subjective or objective [1]. The basic task involved in this is examining adjectives in sentences. Sometimes, the text may be thought of as objective when it has no polarity at all. The second task is to analyze the subjective text in order to determine which of two opposing sentiment polarities it has [1]. This polarity has different strength which differs from one opinion to another. One example of this is that user reviews need to be classified as positive or negative with regard to the target; this shows binary polarity. The work will be more difficult when the polarity is expanded to include more than two items, such as Sorry, Hugs, You Rock, Wow [4], or emotion types, such as happiness, sadness anger, horror and so on [5]. Here, the task becomes a multi-class classification problem.

Another type of sentiment analysis is one that deals with the sentiment target or the discovery of the sentiment target. Most work that has been done in the sentiment analysis field relates to finding sentiments regarding a general topic or target, such as user reviews on a movie or product. In these examples, it is easy to determine the topic, as there is an assumption that the review talks about the specific product. Conversely, it is more difficult in the case of an unknown target, such as with feature-based sentiment analysis. It is difficult to establish what features of the product a user has written about, and then to determine the user's opinion of it. Therefore, an exploration is first made in this situation to establish what features a user has written about by using feature extraction approaches [6], then the sentiment or opinion of these features is determined.

3. Sentiment Analysis Methodologies

A large range of approaches and techniques are used to investigate the problem of sentiment analysis. Most of these approaches are built to deal with the English language as it is the dominant language of science. However, this should not stop researchers from building techniques that work with other languages, such as Chinese, Korean, Japanese and Arabic. This section describes the concepts and research that are used for sentiment analysis in English.

3.1. Technique Types of Sentiment Analysis

Generally, there are two types of techniques that are used in the field of sentiment analysis. These techniques are based on semantic orientation (unsupervised methods) and labeled data (supervised methods) approaches. The following sections shed some light on these techniques.

3.1.1.Semantic-based Approach for Sentiment Analysis

Using semantic orientation for sentiment analysis is considered to be an unsupervised learning method. This technique requires no prior training data in order to find the polarity. It uses statistical inferences to learn from the data and to measure how the words incline toward either positive or negative. Moreover, this type of technique is concerned with linguistically investigating the text to identify the semantic orientation at word, phrase or sentence level [7].

In sentiment analysis, the word usually has one of two types of semantic orientation, either positive or negative. There are different dimensions in the semantic orientation of the word: direction and intensity [8]. The direction dimension refers to the actual polarity of the words, either positive or negative. Regarding the other dimension, the intensity indicates how strong this polarity is. For instance, a review on a product site could indicate a more negative attitude than another negative review. The semantic orientation not only includes the full meaning word (adjective or adverb) but also uses other parts of speech, such as conjunctive words, to improve training in supervised techniques [9]. In the case of the conjunctive "and", both its parts (adjectives) would have the same polarity orientation. On the other hand, the "but" indicates the opposite meaning and polarity.

The research conducted by [10] is considered a significant study. He classified the review using the semantic orientation of words. In this research, the POS is used to extract the phrases that have at least either one adverb or one adjective. The technique, called semantic orientation using point-wise mutual information (SO-PMI), is used to calculate the semantic orientation for the selected phrase. SO-PMI calculates the proximity of the phrase to a predefined set of anchor words. By averaging all values of SO-PMI for all phrases, the overall semantic orientation or polarity of each review is determined. 84% accuracy was achieved for automobile reviews using this method, and 66% for movie reviews.

Other methods in this area depend on the sentiment lexicon to determine the semantic orientation of the review. A lexicon, sometimes called a dictionary, is a list of subjective words annotated with their polarity. The issue here is how the lexicon could be built. In order to determine this, the seeds of positive and negative words (a basic lexicon) are found. A set of pre-existing language sources, such as dictionaries and WordNet [11], is then used to propagate the seeds within it. This works by searching for the synonyms and antonyms of the words in the basic lexicon. If the word found is a synonym for the word in the basic lexicon, then it is added as positive in a lexicon, while it is added to the negative polarity if it is an antonym. This method was used in the work of T. Wilson and his colleagues [12] for the English language.

3.1.2. Supervised Learning-based Approaches for Sentiment Analysis

This approach usually starts with a set of training data. The data should be chosen and categorized properly in order to achieve good prediction results. If not, the data requires a manual effort from the annotator to annotate the data with its subjectivity and polarity. Sometimes, the websites that contain user reviews have ratings along with the reviews. A set of data like this is called a corpus. Next, some of the features that were explained earlier in Section 3 are chosen to represent the text (the review). The next step is to train a classifier on the corpus, and the performance of the classifier is then evaluated on the testing data. This process is usually repeated in an iterative manner if the initial performance is weak. During this repetition, some of the features may be fine-tuned or some preprocessing carried out, including word stemming and the removal of stop words.

The sentiment analysis classification problem can be solved by supervised techniques in the machine learning field. The supervised techniques are data-driven approaches, as data needs to be labeled as an input to the technique. The classifier can then build a model from this training data and apply it to new/unseen/test data in order to determine its class. The crucial factor in supervised methods is the quality and quantity of the trained data. This has a major impact on the accuracy of each technique in this field. A number of learning algorithms have been used in sentiment analysis. The following section will describe research that uses these algorithms in both English and Arabic.

Three different machine learning approaches were investigated by B. Pangand his colleagues [13]. They employed Naïve Bayes, support vector machine SVM and maximum entropy classification, and inferred that the machine learning techniques are better than human baselines for sentiment classification. In addition, the results show that the performance of the SVM was better than other classifiers. The SVM is used with specific features, including unigrams and lemmatized unigrams [14]. They showed that their approach outperformed other approaches that did not use computations for these features. A combination of classifiers was used by R. Prabowo and M. Thelwall [15]. The basic idea in this research was to build hybrid classification. In this process, the document that is not classified by one classifier is sent to the next classifier, until either the document is classified or there are no more classifiers. General inquirer based classifiers GIBC, rule-based classifiers RBC, statistics-based classifiers SBC and SVM are used in this method. It was discovered that SVM and SBC improved the performance of the method. Finally, when comparing supervised and unsupervised approaches, P. Chaovalit and L. Zhou [16] showed that supervised methods achieved 84.49% accuracy for three-fold cross validations and 77% for unsupervised methods with movie reviews.

3.2. Corpus in Sentiment Analysis

The first step in supervised learning approaches in the field of sentiment analysis is the obtaining of a good corpus. Selecting or building a corpus is therefore not an easy job, due to a number of different dimensions. These include the protocol that was used during the annotation process and decisions about what should be annotated [17]. Generally, the corpora carried out in the sentiment analysis field can be categorized depending on the level of document granularity that has been annotated. These granularities include document, sentence, phrase and word level.

At the document level, the whole document is assigned a label that shows the general sentiment of this document. As a result, every word in this document is assigned the same polarity. One of the studies carried out in English in this category was conducted by B. Pang and his colleagues [13]. This corpus contains film reviews from the website IMDb [40], and was released to the research community in 2002. Similarly, A. Abbasiand his colleagues [18], built their corpus in multiple languages, including English and Arabic. The corpus was built from the messages of forum users in the USA and the Middle East.

The sentence level is used to annotate the document at a more granular level. In 2005, B. Pang and L. Lee [19] built a corpus at sentence level from film reviews. The annotation process here was the use of a fine-grained scale rather than a binary scale that showed only the subjective sentences. The fine-grained scale range was from 0 to 5, simulating the star system employed by users of the website, so there was no need for humans to annotate this kind of data. The annotation process was done manually, with 500 positive messages and 500 negative ones annotated.

At the phrase level, C. Potts and F. Schwarz in [20] used n-gram variations to build a phrase-level corpus from documents collected from the Amazon, TripAdvisor and MyPrice websites. Around 700,000 documents were each annotated at different levels of n-gram: tri-, bi- and uni-gram levels of granularity. Each phrase here had a rating ranging from 1 to 5 that reflected the same review score on the respective websites.

3.3. Common Features in Sentiment Analysis

Sentiment analysis is considered a classification problem that can be solved by using the machine learning concept. Machine learning provides many algorithms that work for classification, but the challenge of finding a sentiment in a text is determining the best features to be used. The following sections reveal the common features that are used in sentiment analysis.

Term frequency is the measurement of how many times a specific term is repeated in a document. This has long been emphasized in traditional information retrieval systems. The term presence shows the existence of the term in the document in a binary mode. The document model here shows that term presence is 1 if the term appears at least once in a document, and 0 if not. The term presence model is used in [13] and shows improvement compared with the term frequency model. Famous model is one that uses term frequency and decreases the effect of the high frequency term by using the inverse document frequency (TF-IDF) [21]. The IDF determines whether the term is common or rare across all the documents.

In sentiment analysis, it is important to find the adjectives, as these are good indicators conveying the sentiment orientation in the text [22]. Using the part- of-speech (POS) tagging system decreases the ambiguity of the word [23]. When a word is annotated with its POS tag, this helps to increase the NLP system's confidence in its actual meaning. The POS will help to determine the correct meaning of the word. P. Turney in [10] used the POS feature for adjectives and adverbs in order to obtain the sentiment orientation at document level.

Some other features, such as the style of the text, may contribute to the sentiment orientation of the text. The stylistic features include the length of the sentences, the length of the words, special characters, richness of words, etc. Some research has shown the effect of using the length of sentences as a feature in sentiment analysis [24]. In addition, Abbasiand his colleagues in [18] investigated more than one stylistic feature in multi-language sentiment analysis, including English and Arabic. They found, for example, that in both languages the positive sentiment text is shorter than the negative sentiment text in terms of the total number of characters. They also found that using stylistic features in addition to other features increases the performance of sentiment analysis in web forum discourse [18].

Negation plays a role in sentiment analysis. Negation words can reverse the meaning of a sentence, as a result of which the sentiment orientation should be changed. For example, (انا احب هذه القصة /AnA Ahb hðh h AlqSħ/ "I like this story") attributes a positive sentiment to the story, whereas (انا لا احب هذه القصة /AnA lA Ahb hðh h\; AlqSħ/ "I do not like this story") negates the meaning as well as indicating negative sentiment. These two sentences are very similar, the difference between them being only one word. However, not using negation will negate or reverse the sentiment orientation. For instance, in the sentence (لا عجب ان الجميع يحب هذه القصة /lA ςjb An Aljmyς yHb hðh AlqSħ/ "No wonder everyone loves this story"), the (لا/ lA/ "No") word here does not negate the meaning. In order to deal with this situation, a specific POS tag pattern is used to identify the negations relevant to the sentiment polarity or phrase [24]. Therefore, sentiment analysis should pay careful attention to negation words.

4. Arabic Language and Its Challenges in Natural Language Processing

The Arabic language is one of the most widely used languages, spoken and written by more than 220 million people in over 57 counties [25]. There are three main types of Arabic which are Classical Arabic CA, Modern Standard Arabic MSA, and Dialect Arabic DA [2, 26]. CA is the oldest version of Arabic which is used in the earliest age of Arabic nation, whereas MSA is the formal Arabic language which is used nowadays in education, books, newspapers, media, and even as the official language of Arabic countries. DA is a kind of colloquial language which differs from region to region in Arab countries. There are similarities between MSA and CA since MSA is based on the same syntax, morphology of CA [27], but there are many differences between MSA and DA.

Arabic is comprised of 28 letters. It is a cursive language, in which words consist of cursive Arabic letters connected to one another. Arabic letters have different shapes depending on their position in the word. Unlike English, which has dedicated letters to represent short vowels, Arabic has diacritics that determine the pronunciation or the sound of the letter. In addition, the writing in Arabic runs from right to left. Like other languages, such as Chinese, Japanese and Korean, Arabic has no capitalization. There are two types of Arabic sentences: nominal and verbal; these are determined by the part-of-speech of the first word in the sentence. A nominal sentence has no verb; rather, it is formed of a subject and a predicate. These vary from very simple forms, which consist only of nouns and adjectives, to more complicated ones in which the subject is a compound of two words and the predicate is another sentence [27]. On the other hand, verbal sentences start with a verb and follow different structures and orders. The standard structure of the verbal sentence is verb-subject-object (VSO) [27].

The Arabic language is highly inflectional and derivable. Arabic has a small number of roots, but this increases its complexity. The agglutinative feature of the word structure adds considerable difficulty to the language morphology [27]. Arabic words may work with three types of affixes: prefixes, infixes, and suffixes. Affixes may be one letter long or a combination of multiple letters. In addition to their complex nature, the level of ambiguity of Arabic morphemes is notable. Determining whether a letter is an affix or part of the stem is not an easy task, especially when there is an absence of short vowels. These characteristics affect the NLP tools that deal with Arabic, such as the part-of-the- speech tagger, morphology analyzer, name entity recognition and syntactical parsing. Several studies have been conducted around this.

In terms of Natural Language Processing, other languages, such as English and Chinese, have been the subject of much more investigation, and there are a greater number of tools to help in this field. In the case of Arabic, this field has become increasingly important, and several tools and systems have been developed for different applications [2, 26]. The Arabic language is highly inflectional and derivable, although Arabic has a small number of roots. This increases its complexity. The agglutinative feature of the word structure adds considerable difficulty to the language morphology [27]. Arabic words accept three types of affixes: prefixes, infixes and suffixes. Affixes may be one letter long or a combination of multiple letters. In addition to their complex nature, the level of ambiguity of Arabic morphemes is notable. Determining whether a letter is an affix or part of the stem is not an easy task, especially when there is an absence of short vowels. For instance, (وفي- /wfy/ - “Wafi”, and “in”, or “faithful”) could be either a proper noun, a conjunction followed by a preposition, or even adjective.

The Arabic language these days suffers from diglossia phenomena. The diglossia phenomena manifests when speakers of the same language use two different types of language side by side. In Arab countries, people tend to use MSA in formal communication, media and newspapers, while using the dialect language in their informal speech or even in writing which tends to occur in text messages and over social networks. MSA has specific rules that govern it, while the dialects have different rules and specific structure depending on the type of dialect. This will therefore affect the building of Arabic Natural Language Processing ANLP tools, as the tools that are built for MSA will not work with equal efficiency for DA.

The absence of rigid and strict rules in adding punctuation in MSA text makes it very hard to identify the sentence boundaries [26]. This issue is also a significant challenge in DA, as there are no rules governing it. People often used to write whole paragraphs without using punctuation, except for the full stop at the end. Literal conjunctions, such as (و- /w/ - “and”), are used to organize and link the sentences. This challenge has a direct impact on ANLP, especially for the task of Sentiment Analysis SA, and particularly when selecting proper sentences from entire texts.

5. Related works in Arabic Sentiment Analysis

Much of this research has been done in English, as this is the dominant language of science. Recently, a few researchers have concentrated on applying sentiment analysis to other languages, one such language being Arabic. Figure 1 shows the difference between the research that has been conducted in the Arabic and English languages.

This data is collected by using relevant keywords in sentiment analysis field in both languages. The Google Scholar website is used to collect the numbers of research. For a particular keyword, the Google Scholar is used for a specific period. The results that are retrieved are shown in the top page of the Google website result. These results are used in our comparing.

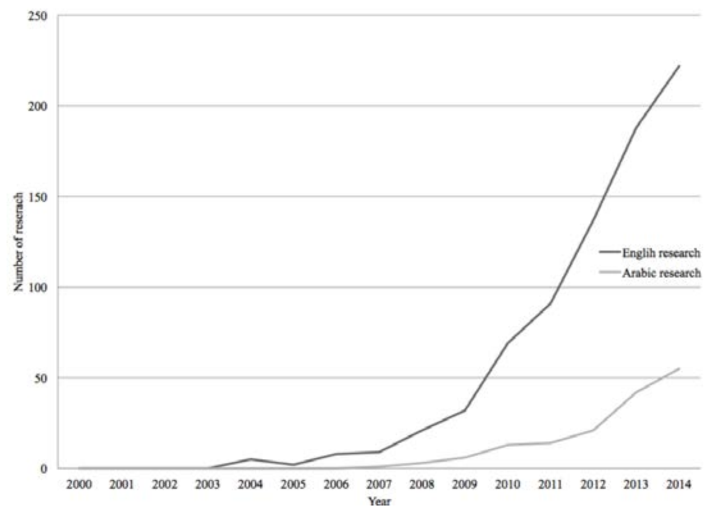

Figure 1: The difference in research that has been conducted in Arabic and English

It is clear that there is a big gap between the work that has been achieved in Arabic and English, Figure 1. This might be due to limitations in the tools or resources of the NLP of Arabic. In addition, it may reveal that Arabic requires special treatment due to its complex nature and structure.

This section summarizes related work that has been done in Arabic sentiment analysis. The summarization is organized into subsections titled to Arabic sentiment corpora, features and methods, and negation.

5.1. Arabic Sentiment Corpora

Arabic sentiment corpora are still in their early stages. Figure 2 illustrates the top ten languages on the internet. These statistics were captured in 2013 according to the Internet World State [28]. This may reveal why most research is conducted in English as well as Chinese; there are plenty of sources in these languages on the internet. However, the Arabic language is considered to be among the top ten languages (fourth position). A small number of research studies have been carried out in this direction. Most researchers in Arabic sentiment analysis built corpus, manually annotating it at either the document or sentence level.

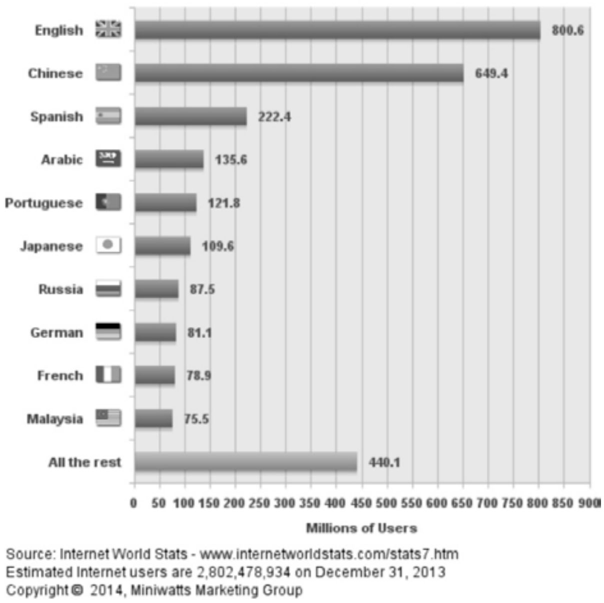

Figure 2: The top ten languages in the internet

M. Elhawary and M. Elfeky in [29] tried to use the same concept to build an Arabic lexicon. They started with over 600 positive words/phrases and over 900 negative words/phrases as seeds. Using an Arabic similarity graph, which they built to compute the polarity score for each word depending on how it was similar to the word in the seeds, they were eventually able to expand the Arabic lexicon.

This method has some limitations, one of which is that it is not necessary for all synonyms to convey the same sentiment orientation. Therefore, the manual annotation has an advantage in building the sentiment lexicon, especially if the annotator is an expert in this field. However, this process is error-prone and time-consuming. A research study in [30] has been conducted to tackle these challenges as well as to deal with informal Arabic (colloquial). A. Al-Subaihin and his colleagues in [30] used the concept of human-based computing to build a game that allows users to annotate words and phrases with their polarity. This game provides each team member (teams consist of two players) with a sentence, and asks them to highlight all the positive and negative words or phrases. The winning team is the one that has the highest agreement between its members. The output of this game will be an informal Arabic lexicon that is used to analyze the sentiment in a sentence.

The Opinion Corpus for Arabic (OCA) [31] (which is the only published corpus) contains 500 movie reviews. They are annotated at the document level. Half the reviews are considered positive and the rest are negative. N. Farra, in [32] built an Arabic corpus and annotated the polarity at the document level. The issue in this corpus was that it was comprised of only a small number of documents (44), of which 27 were positive, 12 negative and five neutral documents.

M. Abdul-Mageedand his colleagues [33] built an Arabic corpus containing 2855 sentences which were collected from the Penn Arabic Treebank [41]. The annotation process was carried out manually by two educated native Arabic people. Each sentence was annotated with objective, subjective-neg (negative polarity), subjective-pos (positive polarity), or subjective-neut (neutral polarity) tags. Further work undertaken to build a multi-genre subjectivity and sentiment corpus for modern standard Arabic is called AWATIF [34]. The domain of this data was taken from a news wire in different domains (400 documents), Wikipedia talk pages (around 5342 sentences), and web forums (around 2532 threads from seven web forums). The annotation was at the sentence level and three different conditions were used to annotate the data: (1) Gold Human with Simple Guidelines (GH-SIMP); (2) Gold Human linguistically-motivated and Genre-nuanced (GHLG); (3) Amazon Mechanical Turk with Simple Guidelines (AMT- SIMP) [35]. In addition, the authors attempted to build a labeled social media corpus for subjectivity and sentiment in the Arabic language in the SAMAR project [36]. The data was collected from four different types of social media. These included Arabic chatting, tweets, Wikipedia Talk, and forums. This corpus was a mix of long and short sentences, as well as MSA and some of DA. They provided stand-off annotations on top of the Arabic Tree Bank ATB [42] part 1 version 3 which is only free for the user who subscribes with the LDC [43] since 2003.

5.2. Features and Methods

Abbasi and his colleagues in [18] proposed a system for sentiment analysis task in a multi-language web forum at document level. The system depends on an Entropy-Weighted Genetic Algorithm (EWGA) to choose the best features, and the SVM with linear kernel for the sentiment classification. Their method tries to find an overlap between language-independent features, including syntactic and stylistic features. The syntactic features include POS only for the English language, not for Arabic. In order to evaluate the performance of their method, the authors measured the accuracy of the classifier by dividing the number of correctly classified documents by the total number of documents. In this case, a more accurate measurement was required to help evaluate the method in both classes. The authors reported that syntactic features achieved a higher result than the stylistic ones. When the two features were employed together using EWGA, the accuracy result increased to 93.6% in the Middle Eastern forum domain.

The work of Rushdi-Salehand his colleagues in [31] focused on investigating two ML classifiers, Naive Bayes and Support Vector Machine, with two different weighting schemes (term frequency and term frequency-inverse document frequency) and three n-gram models. The effect of using the stem of the Arabic work was also investigated with different n-gram models. The authors built their sentiment corpus by collecting around

500 Arabic movie reviews from different websites. They reported an accuracy of 90.6% using the SVM with the tri-gram model and with no stemming for document level classification. In addition, they claimed that there was no big impact of using TF or TF-ID as a weighting scheme, which makes sense because both schemes represent the count of the term over the document. It could be useful to compare the presence of the term versus the term-frequency scheme.

El-Halees in [37] proposed a combined classification approach for document level polarity classification in Arabic. His method applied three different classifiers in a sequential manner: a lexicon-based classifier, a maximum entropy classifier and the K-Nearest Neighbor classifier. The result from one classifier was used as training data for the next. The text was manipulated before using the first classifier by removing the stop words. Some Arabic letters were normalized and some misspelled words corrected. A simple stemmer was used here to generate the stem of the Arabic words and TF-IDF was used as the term-weighting scheme. The F-measure was used as the evaluation metric. The F-measure that was reported in this method was between 75% and 84%, depending on the domain of the data. The average of the F-measure was also calculated; this was 82% for the positive document and 78% for the negative one. The main issue for this study was that there were no more features added to the classifier that could help to increase the performance and accuracy.

Other studies have attempted to investigate the linguistic features of Arabic and to combine these with an ML classifier in order to perform sentiment analysis. One such study tried to analyze the grammatical structure of Arabic [32]. This work attempts to analyze the sentiment at the sentence level first, and then to use the results to analyze the sentiment at the document level. At the sentence level, the researchers compared two different approaches. The first was generalizing the Arabic sentence into a general structure that contains the actor and the action. The second approach used some semantic and stylistic features. The researchers used different classifiers for a different approach. They used the SVM for the grammatical classifier, and obtained an accuracy of 89%, while the J48 decision tree was used with the semantic approaches and achieved an accuracy of 80% when the semantic orientation of the words extracted and assigned manually were used, and 62% when the dictionary was used.

Another work, which investigated the effect of language-independent and Arabic-specific features on the performance of the classifier, was conducted by Abdul-Mageedand his colleagues in [33]. They performed two kinds of sentence level sentiment analysis for two different domains: news and social media domains. The SVM was used to classify both the subjectivity and polarity of the sentences with different features, including N-gram, adjective features and a unique feature, where all words occurring fewer than four times were replaced by the token “UNIQUE”, and MSA morphological features (person, gender and number). By using different stemming and lemmatization settings with different types of independent language and Modern Standard Arabic morphology features, the researchers achieved an F1 result of 72% for subjectivity and 96% for the polarity with stem, morphology setting and ADJ features using the newswire domain. In SAMAR [36], they investigated the effect that the standard features and the genre-specific features had on the subjectivity and sentiment classification of the Arabic social media domain.

5.3. Negation in Arabic Sentiment Analysis

Little work has been undertaken in Arabic in order to address the issue of negation, either in the negation detection problem itself or the effect of negation in sentiment analysis.

Elhawary and Elfeky [29] considered the negation concept in their work. They relied on the Arabic lexicon to calculate the sentiment orientation score of each word or phrase. While the counting process is running, the negated word of the phrase is flipped. There are two main issues here in this work. Firstly, the authors did not mention the Arabic negation words used, stating only that they used around twenty words as negation words. Secondly, there is the issue of how they determined the negated words or phrase that come with the negation word in the sentence. This might affect the process of sentiment analysis, since it has the possibility of changing the polarity (i.e. its polarity type and strength). A further limitation of this work is that the sentiment orientation was calculated depending on the Arabic lexicon, rather than using machine learning to classify the sentiment.

Farra et, al, [32] also considered negation while attempting to capture the sentiment of Arabic text. The negation issue is considered in this work by only counting the frequency of the negation words in the sentence while attempting to build a semantic feature of the sentence depending on Arabic sentiment lexicon. The used features were the frequency of each positive, negative, neutral word, special character and the frequency of the negation words. The authors do not consider the ways in which words might be affected by the negation words. This resulted in a lower accuracy when compared to other methods used by the authors. As in the previous work, the authors here did not mention the list of negation words used. In addition, relying on a simple representation (i.e., frequency counts of negation words or polarity words) would not capture all the semantics and syntax of the sentence that might be useful in sentiment classification.

Hamouda and El-Taher [38] attempted to build a sentiment analyzer for comments on Arabic Facebook news pages. They compared different machine learning algorithms with different features. One of these was dealing with negation in Arabic. They counted only five different negation words, whereas there are many more than these, even without counting negation words in the dialects. They only added the percentage of negation words in either the post or the comment as the feature, without considering the effect of negation on the word or phrase. They claimed that adding negation word features besides the features of all words in the posts and comments gives the best performance. The general issue here is that their proposed method may work only for the domain that they have chosen, which is the posts and the comments in Arabic Facebook news pages. This might, or might not, work with regular Arabic sentiment analysis.

6. Discussion

This section will discuss the limitations and the gaps in the field of Arabic sentiment analysis. At the end, the possible research directions in this field will be presented.

6.1. Arabic Natural Language Processing

Some of challenges and limitations in this field stem from the challenges of the Arabic language itself. As explained earlier, the Arabic language has a rich morphology and high inflection. This affects the natural language processing NLP tools that deal with Arabic, such as the POS tagger, morphology analyzer, name entity recognition NER, and syntactical parsing. Several studies have been done in this direction. The main issue appearing here is that there are a number of well-performing basic NLP tools have been achieved but these tools need more work to be use more easily and could integrated with other programing tools.

Another property of the Arabic language that affects this field is the dichotomy of the Arabic. As mentioned earlier, there are different types of Arabic, including MSA and DA. Most work conducted in NLP for Arabic has been for MSA. The problem is that tools which work well with MSA may not work as well with DA; therefore, there is a need to build a native tool that works especially well with DA [2]. This may also affect the preprocessing approach, such as by removing the stop words and determining the sentence boundaries. In the case of DA, there is a lack of use of punctuation, and there is no capitalization in Arabic. Therefore, the task of determining the sentence boundaries is a crucial and challenging task for NLP in the Arabic language, especially for the task of sentiment analysis.

6.2. Corpus for Sentiment Analysis in Arabic

Most of the research worked with one type of Arabic language, which is Modern Standard Arabic (MSA). Only one work began highlighting and investigating Dialect Arabic (DA) [34]. In addition, there are two problems with this research. The first is the variety of DA used. The study that considers DA only included one form of DA, such as Egyptian Dialect. Each dialect contains different words and expressions that may differ in expressing subjectivity or sentiment orientation. It cannot be ensured that a method that works with Egyptian Dialect would work well with other dialects. The problem that relates to the dialect language is the lack of resources and tools. There are not enough sentiment corpora for the different dialects available to be used. In addition, the Arabic NLP tools that deal with basic NLP tools, such as POS tagger and morphology analyzer, are not yet mature, and are sometimes non-existent for the DA. Therefore, further investigation of the DA is encouraged in subjectivity and sentiment analysis in order to establish which features and ML algorithms work well with DA.

The size and domain of the data sets that are used in subjectivity and sentiment analysis are other issues. Despite some studies reporting high accuracy, this may not always reflect perfection in the proposed method, but may instead be a result of the small size of the dataset used in the experiments. In addition, some of the studies only considered either the news wire or the movie reviews domains. However, what happens if other domains are considered, such as business reviews or even different sub-domains within the main domain, such as different types of news? Moreover, the features or the ML algorithms that work in one domain may not work with the same efficiency in other. It may be useful to use some multi-domains in order to find generalization features and methods that may work with the same efficiency for other types of data domains.

6.3. Techniques Used in Arabic Sentiment Analysis

Table 1 gives a summary of studies that have been done in Arabic sentiment analysis. This table illustrates the approaches that are used in sentiment analysis, the level of sentiment annotation, the corpus domains, the performance that is achieved, and some of advantages and disadvantages for each study. For example, the third study used SVM as machine learning classifier to classify the sentiment into two level either document or sentence in news domain. This study tries to generalize the Arabic sentence into general structure which is considered the most advantage of this study, whereas using a small data in the experiment was the weakness point.

The method used to tackle the problem of how to start classifying the Arabic language is a crucial factor. First, the preprocessing phase for Arabic in order to train the ML classifier plays the main role. Despite this, most of the studies on Arabic sentiment analysis did not explain this phase in detail. Incorrect words, letters with the same shape and effect of the word, such as "أ", "إ" and "ا", and stop words all need to be corrected, normalized or removed. This process should also be undertaken in the case of DA. Secondly, most of the proposed methods in this field used the SVM as the ML classifier with a linear kernel. The Arabic language is recognized to be a highly inflectional and richly morphological language; other classifiers may work better with this language. For example, using a nonlinear kernel with the SVM, or even using the Neural Networks, may lead to better analysis of the sentiment in Arabic, especially in the case of DA, when there is a lack of NLP recourse.

It is clear that most of these works rely on machine learning approaches during the classification process. The most used classifier in these research studies was SVM. However, this may not reduce the power of the other machine learning methods, such as neural networks. L. Chen and H. Chiu in [39] proposed a method to classify sentiment based on neural networks. The neural networks method was trained using three semantic orientation indexes: semantic orientation from association, PMI (point-wise mutual information) and LSI (latent semantic analysis). Therefore, this method might be used with Arabic sentiment analysis.

Table 1: Some of studies that have been done in Arabic sentiment analysis

Studies	Technique	Level	Domain	Pros	Cons
[18]	ML-SVM	Document	Extremist web forums	Use stylistic features Feature selection methods A bilingual system	Domain is very limited More computation cost How they pre-process the Arabic data
[33, 34, 35, 36]	ML-SVM	Sentence	News Social Media	Multi-genre Deploys linguistics-nuanced Involves morphological features Developing various lexicons	Data is limited to the MSA
[32]	ML-SVM	Document Sentence	News	Generalize the sentence structure	Small corpus
[29]	ML-AdaBoost	Document	Business reviews	Builds large-scale lexicon	Do not use Arabic features
[37]	ML	Document	Three domains (Politics, sports, education)	Using more than one ML approaches	Do not use Arabic features

Key: ML: Machine learning, SVM: Support vector machines

While dealing with the negation in Arabic sentiment analysis, most of the works touch the basic idea of how to add and deal with negation during the sentiment analysis processing. The negation tools and styles should be specified during the first step. Previous works either depend on basic negation form or do not mention the negation syntax that they rely on. Moreover, most research in Arabic sentiment analysis does not deal with the issue of negation words while using machine learning algorithm to solve sentiment analysis. They use to using semantic approaches by counting the number of opinioned words instead of using machine learning techniques and flip the score of negated words or counting negation words and adding this to the total score. Therefore, this dissertation tries to come up with a comprehensive method to deal with negation by using machine learning techniques to solve Arabic sentiment classification.

Lastly, the English language has been more investigated in the field of sentiment analysis compared to what has been done in Arabic. It may be useful to benefit from these tools to capture the sentiment in Arabic, but the question that may arise here is “Is the sentiment preserved across different languages during translation process?” This does not imply that there is no need to conduct any further sentiment analysis work for English but getting benefits from what have been achieved in English.

7. Open Research Direction

The field of Arabic sentiment analysis is relatively new. Researchers have put some effort into filling the gaps in this area. Some work has been done to build a data corpus and lexicon for sentiment in Arabic, while other work has concentrated on exploring the techniques used to classify the sentiment in Arabic text. Unlike the English language, in which there have been many studies into sentiment, more work is still required in Arabic in order to build more mature systems for sentiment analysis. This section will discuss some possible work which is represented as having an open direction in this field.

The first direction could be related to the notion of sentiment itself. Most of the work that has been done in Arabic is related to capturing the polarity of the subjective text, either positive or negative. This is acceptable, as work in the Arabic sentiment analysis is still at an early stage. One possible direction here is to classify the text as fine-grained sentiment. Instead of using two polarities to represent the sentiment, a scale range from 0 to 10 can be used to show the sentiment strength in the text that could be used (0 means more negative sentiment, 10 means more positive sentiment). Using two kinds of polarity only captures the type of sentiment in the text, whereas using a fine-grained polarity could capture the type and strength of the sentiment.

Carrying out more work on the dataset for an Arabic sentiment corpus and lexicon is a challenging task and needs more attention, as it is an important part of the sentiment analysis process. Some initiatives have been implemented and discussed in the previous section. As explained, only one of them has been publicly published for the benefit of the research community. The studies that may be involved here may develop in a number of different directions. The domain of the chosen data represents one direction. Most work in Arabic Sentiment Analysis ASA deals with the news or business/movie reviews domains. Other types of data could be also included, such as political discussion or medical domain. In addition, social networks today are a de facto method of communication between people; this could use as a source for building sentiment corpus in this field.

Regarding to language types, most of the work in this field has been carried out on the type of Arabic language known as MSA. Dialect Arabic DA requires more work and investigation. Recently, M. Abdul-Mageed and M. Diab in [35] included this type of language in their work, but they only included one Arabic dialect, the Egyptian dialect. There is large scope for work to be done here in order to build a more sophisticated corpus that includes different Arabic dialects. The last direction for this type of work to go in could include the method of labeling the data in the corpus. Most of the work has been done manually, which costs the consumer time and money. Therefore, a fully semi-automated method is needed to label the data with its sentiment. A. Al-Subaihin in [30] used a kind of human-based game to involve more people in labeling the text with its sentiment while they were gaming. Using a well-known corpus from a well-studied language such as English and translating it to Arabic could be a useful approach to building a quick corpus. Therefore, more studies need to be conducted in this direction in order to improve the viability of this method.

Machine learning ML approaches have a strong role in classifying Arabic sentiment. Most of the studies in Arabic use the SVM, to classify the sentiment in the Arabic text. This may be another direction for researchers to investigate further by exploring other MLs in the field of Arabic sentiment, such as neural networks, as has been done in the English language [39]

8. Conclusion

This article describes the field of sentiment analysis in the Arabic language, and sheds some light on what has been accomplished in the English language as an example of a language that has been well studied for sentiment analysis, compared with what has been investigated in the Arabic language. Moreover, some novel studies and works have been conducted in Arabic sentiment analysis. However, these works are still at an early stage and require more effort from researchers in order to increase the quality and strength of the tools in this field.

This field is still struggling with a number of issues, beginning with the scarcity of linguistic resources for Arabic and ending with the lack of corpora sources for Arabic sentiment and the approaches used to classify the sentiment in the text. The internet provides an unlimited source of material that contains sentiment in different aspects, as well as varieties of Arabic dialect that are used in these sites, such as social networks.

References

[1] Pang, Bo; Lee, Lillian. 2008. Opinion Mining and Sentiment Analysis, Found. Trends Inf. Retr. 2, 1-2, pp. 1--135.

[2] Farghaly, Ali; Shaalan, Khaled. 2009. Arabic Natural Language Processing: Challenges and Solutions. In ACM, 2009, 8, 14:1-14:22

[3] Habash, Nizar; Soudi, Abdelhadi and Buckwalter, Timothy, 2007. On Arabic Transliteration, In Arabic Computational Morphology, pp. 15--22.

[4] Socher, Richard; Pennington, Jeffrey; Huang, Eric H.; Ng, Andrew Y.; Manning, Christopher D., 2011. Semi-supervised recursive autoencoders for predicting sentiment distributions, In Proceedings of the Conference on Empirical Methods in Natural Language Processing (Stroudsburg, PA, USA:

Association for Computational Linguistics, 2011), pp. 151--161.

[5] Subasic, Pero; Huettner, Alison, 2000. Affect analysis of text using fuzzy semantic typing, In Fuzzy Systems, 2000. FUZZ IEEE 2000. The Ninth IEEE International Conference on vol. 2, pp. 647 -652 vol.2.

[6] Popescu, Ana-Maria; Etzioni, Oren, 2005. Extracting product features and opinions from reviews, In Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing (Stroudsburg, PA, USA: Association for Computational Linguistics, 2005), pp. 339--346.

[7] Brooke, Julian, 2009. A semantic approach to automated text sentiment analysis. PhD diss., Simon Fraser University.

[8] Turney, Peter D.; Littman, Michael L., 2003. Measuring praise and criticism: Inference of semantic orientation from association, In ACM Trans. Inf. Syst. pp. 315--346.

[9] Hatzivassiloglou, Vasileios; McKeown, Kathleen R., 1997. Predicting the semantic orientation of adjectives, In Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics (Stroudsburg, PA, USA: Association for Computational Linguistics), pp. 174--181.

[10] Turney, Peter D., 2002. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews, In Proceedings of the 40th Annual Meeting on Association for Computational Linguistics (Stroudsburg, PA, USA: Association for Computational Linguistics), pp. 417--424.

[11] Miller, George A., 1995. WordNet: a lexical database for English, In Commun. ACM, pp. 39--41.

[12] Wilson, Theresa; Wiebe, Janyce; Hoffmann, Paul, 2009. Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis, In Computational linguistics 35, pp. 399--433.

[13] Pang, Bo; Lee, Lillian; Vaithyanathan, Shivakumar, 2002. Thumbs up?: sentiment classification using machine learning techniques, In Proceedings of the ACL-02 conference on Empirical methods in natural language processing - Volume 10 (Stroudsburg, PA, USA: Association for Computational Linguistics), pp. 79--86.

[14] Mullen, Tony; Nigel Collier, 2004. Sentiment analysis using support vector machines with diverse information sources, In Proceedings of Conference on Empirical Methods in Natural Language Processing.

[15] Prabowo, Rudy; Mike Thelwall, 2009. Sentiment analysis: A combined approach, In Journal of Informetrics 3, pp. 143 - 157.

[16] Chaovalit, Pimwadee; Lina Zhou, 2005. Movie Review Mining: a Comparison between Supervised and Unsupervised Classification Approaches, In System Sciences, HICSS '05. Proceedings of the 38th Annual Hawaii International Conference, pp. 112c.

[17] Sorokin, Alexander; Forsyth, David, 2008. Utility data annotation with Amazon Mechanical Turk, In Computer Vision and Pattern Recognition Workshop,. CVPRW '08. IEEE Computer Society Conference, pp. 1 -8.

[18] Abbasi, Ahmed; Chen, Hsinchun; Salem, Arab, 2008. Sentiment analysis in multiple languages: Feature selection for opinion classification in Web forums, In ACM Trans. Inf. Syst. pp. 12:1--12:34.

[19] Pang, Bo; Lee, Lillian, 2005. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales, In ANNUAL MEETING-ASSOCIATION FOR COMPUTATIONAL LINGUISTICS vol. 43, no. 1, pp. 115.

[20] Potts, Christopher; Schwarz, Florian, 2008. Exclamatives and heightened emotion: Extracting pragmatic generalizations from large corpora, Ms., UMass Amherst.

[21] Sparck Jones, Karen, 1988. Document retrieval systems, In Willett, Peter, ed., (London, UK, UK: Taylor Graham Publishing), pp. 132--142.

[22] Benamara, Farah; Cesarano, Carmine; Picariello, Antonio; Reforgiato, Diego; Subrahmanian, V. S., 2007. Sentiment analysis: Adjectives and adverbs are better than adjectives alone, In Proceedings of the International Conference on Weblogs and Social Media (ICWSM).

[23] Wilks, Yorick; Stevenson, Mark, 1998. The grammar of sense: Using part-of-speech tags as a first step in semantic disambiguation, In Nat. Lang. Eng. pp. 135--143.

[24] Na, Jin-Cheon; Haiyang Sui; Christopher Khoo; Syin Chan; Yunyun Zhou, 2004. Effectiveness of Simple Linguistic Processing in Automatic Sentiment Classification of Product Reviews, In Conference of the International Society for Knowledge Organization (ISKO), pp. 49--54.

[25] Lewis, M. Paul, 2009. Ethnologue: Languages of the world vol. 9, In SIL international Dallas, TX

[26] Shaalan, Khaled. 2101. Rule-based Approach in Arabic Natural Language Processing, In the International Journal on Information and Communication Technologies (IJICT) 3, pp. 11--19.

[27] Ryding, Karin C. 2005. A reference grammar of modern standard Arabic. In Cambridge University Press.

[28] I. W. Stats. Internet world users by language. Miniwatts Marketing Group, 2013. http: //www.internetworldstats.com/stats7.htm, accessed 31 March, 2015.

[29] Elhawary, Mohamed; Elfeky, Mohamed, 2010. Mining Arabic Business Reviews, In Data Mining Workshops (ICDMW), 2010 IEEE International Conference, pp. 1108 -1113.

[30] Al-Subaihin, Afnan A.; Al-Khalifa, Hend S.; Al-Salman, AbdulMalik S. 2011. A proposed sentiment analysis tool for modern Arabic using human-based computing, In Proceedings of the 13th International Conference on Information Integration and Web-based Applications and Services (New York, NY, USA: ACM), pp. 543--546.

[31] Rushdi-Saleh, Mohammed; Martıń-Valdivia, M.T.; Ureña-López, L.A.; Perea-Ortega, J.M., 2011. OCA: Opinion corpus for Arabic, In Journal of the American Society for Information Science and Technology 62, pp. 2045--2054.

[32] Farra, Noura; Challita, Elie; Assi, Rawad Abou; Hajj, Hazem, 2010. Sentence-Level and Document-Level Sentiment Mining for Arabic Texts, In Data Mining Workshops (ICDMW), 2010 IEEE International Conference, pp. 1114 -1119.

[33] Abdul-Mageed, Muhammad; Diab, Mona T.; Korayem, Mohammed, 2011. Subjectivity and sentiment analysis of modern standard Arabic, In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers - Volume 2 (Stroudsburg, PA, USA: Association for Computational Linguistics), pp. 587--591.

[34] Abdul-Mageed, Muhammad; Diab, Mona, 2012. AWATIF: A multi-genre corpus for Modern Standard Arabic subjectivity and sentiment analysis", In Proceedings of LREC, Istanbul, Turkey.

[35] Abdul-Mageed, Muhammad; Diab, Mona, 2012a. Toward building a large-scale Arabic sentiment lexicon", In Proceedings of the 6th International Global Word-Net Conference, Matsue, Japan.

[36] Abdul-Mageed, Muhammad; Kübler, S.; Diab, Mona, 2012. SAMAR: A System for Subjectivity and Sentiment Analysis of Arabic Social Media, In WASSA, pp. 19.

[37] El-Halees, Alaa, 2011. Arabic Opinion Mining Using Combined Classification Approach, In Proceeding the International Arab Conference On Information Technology, Azrqa, Jordan.

[38] A. E.-D. A. Hamouda and F. E.-Z. El-Taher. Sentiment analyzer for Arabic comments system. International Journal of Advanced Computer Science and Applications(IJACSA), volume 4(3): pages 99–103, 2013.

[39] Chen, Long-Sheng; Hui-Ju Chiu, 2009. Developing a Neural Network based Index for Sentiment Classification, In International MultiConference of Engineers and Scientists, Hong Kong vol. 1, (, 2009).

[40] Website, www.imdb.com, last date accessed 1\3\2016.

[41] Website, http://www.ircs.upenn.edu/arabic, last date accessed 1\3\2016.

[42] Website, http://www.ldc.upenn.edu/Catalog/catalogEntry.jsp?catalogId=LDC2005T02, last date accessed 1\3\2016.

[43] Website, http://www.ldc.upenn.edu/, last date accessed 1\3\2016.