Classifying Sentiment of Dialectal Arabic Reviews: A Semi-Supervised Approach

Omar Al-Harbi
Computer and Information Department, Jazan University, Saudi Arabia

Abstract: *Arab Internet users tend to use dialectical words to express how they feel about products, services, and places. Although, dialects in Arabic derived from the formal Arabic language, it differs in several aspects. In general, Arabic sentiment analysis recently attracted lots of researchers' attention. A considerable amount of research has been conducted in Modern Standard Arabic (MSA), but little work has focused on dialectal Arabic. The presence of the dialect in the Arabic texts made Arabic sentiment analysis is a challenging issue, due to it usually does not follow specific rules in writing or speaking system. In this paper, we implement a semi-supervised approach for sentiment polarity classification of dialectal reviews with the presence of Modern Standard Arabic (MSA). We combined dialectal sentiment lexicon with four classifying learning algorithm to perform the polarity classification, namely Support Vector Machines (SVM), Naïve Bayes (NB), Random Forest, and K-Nearest Neighbor (K-NN). To select the features with which the classifiers can perform the best, we used three feature evaluation methods, namely, Correlation-based Feature Selection, Principal Components Analysis, and SVM Feature Evaluation. In the experiment, we applied the approach to a data set which was manually collected. The experimental results show that the approach yielded the highest classification accuracy using SVM algorithm with 92.3 %.*

Keywords: *Arabic sentiment analysis, Opinion mining, Dialectal sentiment analysis, Dialectal lexicon, Dialectal Arabic processing.*

Received January 8, 2017; accepted March 6, 2018

1. Introduction

The improvement in communication and social media generated an enormous amount of valuable data on the web. This data can be utilized to build many useful technology services such as automatic sentiment analysis, question answering systems, etc. Measuring the satisfaction and obtaining the feedback automatically from users have always been the concern of companies that offer services or products to make decisions that would improve their business. Several years ago, this process was impossible, but with the advancement in web 2.0 platforms such as forums, blogs, and social media including their extensive data made that possible. Creating systems that can automatically extract opinionated phrases from unstructured texts like posts, comments, tweets, and reviews is the concern of the sentiment analysis field. Sentiment analysis or opinion mining field is a task of Natural Language Processing (NLP) [27]. This task concerns with subjective information detection in textual information.

The research in English language sentiment analysis has achieved considerable progress, whereas it is still limited in other languages such as Arabic [7]. The complexity of Arabic language and the lack of linguistics resources made the task of Arabic sentiment analysis even harder. The Arabic language requires advanced pre-processing methods and various linguistics recourses, due to its morphological complexity and diversity of the dialects [5, 26].

In general, the Arabic language has three categories [17]: classical Arabic found in Qur'an and religious scripts, Modern Standard Arabic (MSA) used and understood all over the Arabic world, and dialectal Arabic which is usually spoken not written and varies according to the regions. As MSA is the most used language in writing and formal speeches, it attracted computational linguistics researchers to create morphological and syntactic tools that would handle Arabic NLP tasks.

However, in social platforms Arab users usually tend to use their dialects alongside MSA form. Arabic dialects derived from MSA and classical language, and it varies according to some circumstances such as region. In the Arab world, there are several dominant dialects, namely Egyptian, Maghrebi, Levantine, and Gulf. Although these dialects descended from the Arabic language, they are considered distinct languages [35]. Recently, dialectal Arabic has been widely used in writing on social networks, forums, and blogs. Thus, NLP task such as sentiment analysis would be more challenging.

Arabic sentiment analysis with the existence of dialects has become more challenging process, and new processing methods with extending linguistics recourses are highly required. In literature there is few researchers concerned with dialectical Arabic sentiment analysis such as [1, 13, 16, 35]. Most researchers targeted Egyptian dialect since it is the most widely spoken dialect in the Middle East by more than 80 million people. In this research, we concern with reviews written with Jordanian dialect.

Jordanian dialect is spoken by the people of the Kingdom of Jordan, and it belongs to Levantine Arabic. Jordanian dialect has the Semitic language structure, and it is lexically influenced by languages like Turkish, English, and French [34]. To the best of the author knowledge, there is no much research concerned with the Jordanian dialect in sentiment analysis other than [2, 13, 14]. Furthermore, there are no Jordanian dialect resources available publically for sentiment research purposes. Thus, we present an implementation of semi-supervised approach for sentiment analysis of Jordanian dialectal reviews. This approach uses learning classifiers combined with dialectal sentiment lexicon and other dialectal linguistic resources such as dialectal compound phrases, contrary words, and negation words. In this paper, we also present a comparison of features evaluation methods, and their effect on the performance of most well-performed learning classifiers such as Support Vector Machine (SVM), Naive Bayes (NB), Random Forest, and K-Nearest Neighbour (K-NN).

The paper is conducted as follows. Section 2 provides some background on sentiment analysis and The Arabic language and the dialects. Section 3 presents related work. Section 4 introduces the methodology and experiment setting. Section 5 discusses experimentations and results. Finally, Section 6 presents the conclusions of this work.

2. Background

2.1. Sentiment Analysis

Sentiment analysis, also called opinion mining, is defined in [28] as an "interdisciplinary field that analyzes people's opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes."

The core function of sentiment analysis area is assigning a label of positive, negative, or neutral to opinionated words, phrases, or documents. In general, sentiment analysis has been investigated at three levels, namely: Document level, Sentence level, and Entity and Aspect level [28]. Sentiment classification can be performed based on two primary methods [37]. Firstly, a statistical or machine-learning method or can also be described as a supervised approach, which uses in polarity classification machine learning algorithms such as SVM, NB, Decision Trees, etc. Secondly, lexicon-based method or can also be described as an unsupervised approach, which exploits lexicons, dictionaries, and other linguistics resources and rules to classify the polarity.

However, sentiment analysis has many challenging problems that would make it an arduous task, most of the problems reported in [9, 28, 29]. Likewise, [15] highlighted major problems of sentiment analysis in Arabic languages such as the existence of dialects, the lack of Arabic dialects resources and tools, the limitation of Arabic sentiment lexicons, using compound phrases and idioms, etc.

2.2. Arabic Language and Dialects

The Arabic language is used by about 325 million native speakers to daily communication [36]. It is also one of the languages in United Nations as are English, and French. The Arabic script is the second most familiar script in the world after Latin. It is used in Arabic and other languages such as Ottoman Turkish, Persian, Urdu, Afghan, and Malay [30]. However, the Arabic language has a morphologically complex style that has a high inflectional and derivational nature [15]. In Arabic language, MSA is the most common and understood from all over Arabic world, and used in books, newspapers, news, formal speeches, subtitles, etc. The MSA derived from the classical Arabic, and they have several features in common. However, they are treated separately and have differences in aspects such as lexicon, stylistics, and certain innovations on the periphery [22].

Arabic dialects are also rooted in classical Arabic and MSA, and the script is the same. There is a great variety in Arabic dialects among the Arab countries or even different regions at the same country. For example, there are several ways in Arabic to say " ماذا تريد؟" "What do you want?" in Jordanian "أيش بدك" "aysh bedk", in Saudi "وش تبغى" "wsh tabgha", in Egyptian "عايز ايه" "ayz ayh". Obviously, there are no standard rules for dialects at the levels of morphology, phonology, syntactic, and lexicon. Shaalan *et al.* [35] pointed out that the differences between dialects and MSA because behaviours such as replacing characters and change the pronunciation or the style of writing of nouns, verbs, and pronouns. Consequently, new dialectal words will continue appearing, and the gap between MSA and dialects will increase.

The Jordanian dialect is spoken by more than 10 million[1]. According to [11], the Jordanian dialects has three categories. First, the urban dialect which has emerged as a result of internal and external migrations to the main cities. Second, the rural dialect which is often spoken in villages and small cities, and it has two categories; Horan dialect which is used in the area north and west Amman, and Moab dialect which is used in the area of South Amman. Third, Bedouin dialect which is spoken by Jordanian Bedouins who live in the desert, and is not common in the urban and rural regions. Table 1 shows an example of how the Jordanian dialect varies in a sentence like "ما خطبه؟" which means "What is wrong with him?".

[1]http://dosweb.dos.gov.jo/ar/

Table 1. Categories of Jordanian dialects.

Jordaian Dialect	Sentence	Buckwalter
Urban Dialect	ماله هاد	mAlh hAd
Rural dialect	مالو هاظ	mAlw hAZ
Bedouin	علامو هاذ	ElAmw hA*

The migration has played a significant role in the formation of Jordanian dialect. Since 1984, Jordan has received a considerable amount of Palestinian refugees who settled all over the region. The contact between Palestinians and Jordanians has created new and complex patterns of dialects [31]. Furthermore, a flood of Syrian refugees recently was accepted, that made Jordan dialect observably propagated. Based on the introduced facts, Jordanian dialect continuously adds new suffixes, prefixes, and clitics that would generate new words, stop-words, contrary words, and negation words (e.g., مش، مهوش, منو, خنروح, خرمان).

In fact, the dynamic nature of dialect would create complexity in developing sentiment recourses such as lexicons, annotated corpora, and parsers. For example, people use new and different ways to express their sentiment such as transliterated English like (لول, نايس كيوت) which means (LoL, nice, cute), and newly created compound phrases like (سعرو فيه) which mean (worthy). Because such challenges, processing Arabic dialects in sentiment analysis is difficult, and most researchers would rather deal with MSA texts, since MSA was robustly researched and have a considerable amount of resources.

3. Related Work

Processing Arabic dialects in sentiment analysis is relatively a new area of research. In this section, we present some previous work that addressed the issue of Arabic dialects in sentiment analysis. However, most of the studies in the literature of Arabic sentiment analysis focused on MSA form, due to the lack of resources and tools in Arabic dialects [15]. In this section we shed light on some researches related to our work such as [2, 4, 6, 13, 14, 16, 23, 32, 33].

Duwairi [14] introduced a framework for sentiment analysis of Arabic tweets with the presence of Jordanian dialect. The approach utilizes machine learning classifier and dialect lexicon which maps dialectical words into their corresponding MSA words. 22550 tweets were collected using Twitter API and annotated using a crowd-sourcing tool. In this work, utilizing the dialectal lexicon achieved a slight improvement. Two classifiers were used to determine the polarity, namely: NB and SVM, the F-measure of the two classifiers was 87.6% and 86.7% respectively.

Abdulla *et al.* [2] presented a lexicon-based approach for analyzing opinions written in both MSA and Jordanian dialect. The lexicon size was 3479 words, and the dataset composed of 2000 tweet were collected and manually annotated. For feature extraction, they used unigram technique, and then they used an aggregation tool to calculate the weights of tweets to generate the polarity. They performed a comparison between lexicon-based and corpus-based approaches; as noticed from the results corpus-based approach remarkably outperformed the lexicon-based approach. The final reported accuracy of lexicon-based approach was 59.6%.

Ibrahim *et al.* [23] used a semi-supervised approach for sentiment analysis of MSA and Egyptian dialect. They introduced a high coverage Arabic sentiment lexicon with 5244 terms, and a lexicon of idioms/saying phrases with 12785 phrases. Regarding feature selection, they extracted different linguistic features to improve the classification process. For classification, they used the SVM technique. Their dataset consists of 2000 statement divided into 1000 tweet and 1000 microblogging reviews. The reported accuracy of the SVM classifier was 95%.

The work of Mourad and Darwish [32] focused on Subjectivity and Sentiment Analysis (SAA) on Arabic news articles and dialectal Arabic microblogs from Twitter. A random graph walk approach was employed to expand the Arabic SSA lexicon using Arabic-English phrase tables. They used two classifiers in the experiments, the NB and SVM classifiers with features such as stem-level features, sentence-level features, and positive-negative emoticons. The accuracy was 80% for news domain and 72.5% for tweets.

Azmi and Alzanin [6] introduced Aara' which is a mining system for public comments written in Saudi dialect. They employed the Naïve Bayes algorithm with a revised n-gram approach for classification. The dataset consists of 815 comments which were gathered manually from online newspapers, and then split into a training set and testing set. The accuracy of the system was 82%.

Al-Subaihin and Al-Khalifa [4] presented an unsupervised technique for extracting sentiments from informal restaurants reviews. In this work, human interaction is a major component, to annotate the text in an entertaining way. They used two approaches to determine the polarity, namely: sentimental tag patterns with precision 56.14% and sentimental majority approach with precision 60.5%.

Finally, Abdul-Maged *et al.* [1] presented SAMAR for subjectivity and sentiment analysis for Arabic social media reviews. In this work, they considered both MSA and Arabic dialects. In this work, different features were used include author information, stemming, POS tagging, dialect and morphology features. For classification, they used SVM classifier over a variety datasets. Concerning dialectal Arabic, they noticed that the presence of dialectal tweets would affect the SSA negatively since the most tweets are subjective and negative in sentiment. The highest accuracy reported through the dialect-specific sentiment experiments was 73.49%.

As can be noticed from the related work, different approaches, methods, resources, and language have been utilized to analyze different Arabic dialects. Unfortunately, we found only a few research works that concern with Jordanian dialect. We implement a semi-supervised approach to determine the polarity of Jordanian dialectal reviews. In this work, dialectal lexicons were built and combined with different machine learning algorithms (NB, SVM, Random Forest, and K-NN) to find the best classification model. We also investigated different features evaluation methods to improve the classification process.

4. Methodology and Experiment Setting

This section describes the methodology and material used in our work. In this work, we used a semi-supervised approach by combining machine learning classifiers with a dialectal lexicon to classify the polarity.

4.1. Corpus

To learn the classifiers, an annotated training corpus is required. To the best of our knowledge, there is no publicly available corpus for Jordanian dialect. Thus, we manually built our own corpus consisting 2500 reviews of which 1450 were positive, and 1050 were negative. The data was collected from JEERAN/Jordan[2] website which is a platform of users' reviews about places, services, and products in Jordan. The reviews include various domains (restaurants, shopping, fashion, education, entertainment, hotels, motors, and tourism). The reviews are mostly written by reviewers from the public, so it may contain dialectal and MSA terms, and also it can be short or long reviews. Two Jordanian native speakers annotated the polarity of the reviews, and a good agreement was reflected. In this work, only the positive and negative reviews were considered, while the reviews such as neutral, sarcastic, and uncertain have been disregarded in this work.

4.2. Sentiment Lexicons

In this work, we built a lexicon consist of 3400 opinionated term (adjectives, adverbs, and verbs). We manually extracted from our corpus all opinionated terms including dialectal and MSA, and then stored in the lexicon. Furthermore, we built a lexicon contains 580 compound phrases that may have sentiment, and their existence combined together indicates to positive or negative reviews. Table 2 shows examples of dialectal compound phrases and individual words. Words from other dialects and English transliterations were included in the lexicon like (نايس,لايك,جنتل).

[2]http://jo.jeeran.com/en/amman/

Table 2. Sample of the dialectal words and compound phrases.

Dialectal Words & Compound phrases	Corresponding MSA	Polarity	Buckwalter Transliteration	Gloss
مناح	جيدون	Positive	mnAH	They are well
عجئة	إزدحام	Negative	Ej}p	Crowded
أنفلمت	خُدعت	Negative	>nflmt	I have been deceived
سعرو فيه	يستحق	Positive	sErw fyh	Worthy
بطلوع الروح	بصعوبة	Negative	bTlwE AlrwH	Hardly

4.3. Pre-processing

In this phase, pre-processing included correcting misspellings and removing repeated letters in words. We also removed punctuations, numerals, English words, and elongation. Emoticons are also removed because its usage was rare in the collected reviews. Next, a normalization process was applied to particular letters, for example the letters (أ, إ, آ) were converted to (ا), the letters (ى, ئ) were converted to (ي), the letter (ة) was converted to (ه), and finally the letter (ؤ) was converted to (و). The problem of the conjunctive particle (WA, و) was handled by applying a naïve algorithm that simply removes the (و) from the beginning of any word containing more than three letters.

4.4. Features Identification

Feature identification is a fundamental process prior to applying a learning classification algorithm. In this process, the data is transformed into dimensions of features describing the content. The effectiveness of identifying features plays an essential role in obtaining high performance. In this work, we investigated the following 9 features:

1. Positive Words Number (PWN): This feature represents the total number of positive words in the review. To extract this feature and the next feature, we built a dialectal sentiment lexicon which introduced in section 4.1.
2. Negative Words Number (NWN): This feature represents the total number of negative words in the review. We developed a simple algorithm to extract the feature NWN and PWN from the reviews.
3. Negation Words Number (NgWN): In dialects, negation can be expressed in different ways such as (مو,مش,فش,مفيش). Thus, dialectal and MSA negation words have been collected and stored in a negation list. The negation words can change the polarity of any sentiment word to the opposite. To handle this problem, we developed an algorithm that can change the polarity of sentiment words that follow any negation word. Wherever the negation word is found in the review, the algorithm will search any sentiment word within a scope of only the three following words, and then the polarity will be

reversed. After that, the appearances of negation words in the review will be counted and provided as NgWN feature.

4. Contrary Words Appearance (CWA): The presence of contrary words such as (بس, لكن, لاكن) in a review would make the polarity classification challenging task, where the review may start with a particular sentiment and somewhere at the review; these words are used to reverse the sentiment. For example, " قعدة الترس بهذا الكافيه حلوة, بس اسعارو نار والأرجيلة مش ولا بد" that means "Sitting on the terrace of this cafe is beautiful, but it is very expensive and the Shisha in not that good." As noticed, the review started with a positive sentiment, and when the contrary word (بس) was used the sentiment of the review became negative. For the feature of the contrary words, we only considered the presence of these words in the review, and that means it is a binary-valued feature vector in which the contrary words either appear and given the value (1) or do not appear and given the value (-1).
5. Positive Compound Phrase (PCP): In dialects, compounding phrases is a common way to describe emotions and opinions. However, compound phrases may vary from dialect to another. Compound phrases mean that two or more non-opinionated terms together can hold a particular sentiment such as (يكثر خيرهم, قول وفعل, همهم رضاك). To extract this feature from the dataset, we used a lexicon of positive and negative compound phrases containing 580 phrases. This feature is also a binary-valued, where the (1) would be given if it appeared, and the (-1) would be given if it did not appear.
6. Negative Compound Phrases (NCP): This feature represents the appearance of negative compound phrases in the reviews such as (كثير عليهم, فوق هذا كله). Compound phrases also can refer to the idioms like (على عينك يا تاجر), and supplications like (حسبي الله ونعم الوكيل, منك لله, اتقوا الله). This feature is also a binary-valued, where the (1) would be given if it appeared, and the (-1) would be given if it did not appear.
7. Positive Words Positions (PWP): This feature represents the positions of positive words, where this feature can play a useful role in classifying the polarity of reviews. A numeric value has been assigned to the reviews representing the sum of all positions' values of positive words in the review.
8. Negative Words Positions (NWP): This feature also represents the positions of negative words in the reviews. The feature is represented as a numeric value which is the sum of all positions' values of negative words in the review.
9. Review Length (RL): This feature represents the length of the reviews. Each word in the review will be counted, and then the sum of the words will be represented as a numeric value.

Although, there are only 9 features, we used three of the most commonly used feature evaluation methods to select the best features subset, namely: Correlation-based Feature Selection (CFS) [21], Principal Components Analysis [25], and SVM Feature Evaluation [19]. Using these feature evaluators can potentially improve the classifiers' performance, and help in data understanding [20]. In our work, we compared the effectiveness of the evaluation methods in selecting the feature subset with which the classifier may yield the best accuracy.

4.5. Polarity Classification

The goal of classification is to categorize input data into predefined classes. In this work, we have two classes; they are positive and negative. Next step after transforming the data into feature space is selecting the suitable learning classifier. Therefore, in our work, we examined four of the most robust an accurate classifiers that have been used by researchers in data mining [38], namely, SVM [12], Random Forest [8], NB [24], and K-NN [3]. The classifiers represent diverse approaches to learning, and their behavior is suited to the dataset and the vector representation.

5. Experimental Results and Evaluation

This section presents the experimental results for the classifiers used to classify the reviews to either positive or negative classes. In this experiment, we used four metrics for evaluating the performance of the classifiers, they are: Accuracy, Precision, Recall, and F-Measure. The dataset divided into two sets, the first one is a training set which is 70% of the original set, and consists of 1750 reviews of which 980 were positive and 770 were negative, and the second one is a testing set which is 30% of the original set, and consists of 750 reviews of which 472 were positive and 280 were negative.

To perform this experiment we used Weka software [18]. Weka introduces a set of machine learning algorithms and tools for data mining purposes. To recognize the algorithm that would be the best suited to classify our dataset, we compare the performance of four learning classifying methods. The methods are SVM, RandomForest, NB, and K-NN (where K=9 because it gave the best accuracy). Regarding SVM, we used the package LIBSVM which introduced by [10].

The experiment is performed through three phases. In the first phase, training and testing the classifiers were carried out with examining all proposed nine features and with the default parameters of the classifiers. Table 3 summarizes the results of this phase. In the second phase, to yield better performance, we investigated the optimal features with which the classifiers would obtain the highest performance. The decision of selecting these particular features over the other was made after applying three automatic feature evaluation methods (CFsSubsetEval, PrincipalComponentsEval, and SVMAttributeEval) which are embedded in Weka software. In the third phase, after identifying the classifier with the best performance and best feature subset, we tuned the parameters of the classifiers so that better accuracy might be obtained. Tuning the parameters was done with considering challenging issues may rise such as over-fitting and under-fitting.

Table 3. Results of the classifiers with the nine features.

Classifier	Accuracy (%)	Precision (%)	Recall (%)	F-Measure (%)
SVM	90.1596	90.1	90.2	90.1
NB	80.4521	80.3	80.5	80.3
RandomForest	91.0904	91.1	91.1	91.1
K-NN (K=9)	91.2234	91.2	91.2	91.2

Table 3 shows the performance of the classifiers with all proposed features, and we can notice that the four classifiers are convergent in performance except for NB. K-NN classifier outperformed the other classifiers, and the lowest performance is obtained with the NB classifier. The difference between K-NN and NB is about 11%, while with RandomForest the difference is approximately 0.13%, and with the SVM is approximately 1%.

To improve the classification and obtain a higher accuracy, we selected the optimal features based on the three suggested feature evaluators. These feature evaluators selected and ranked subsets of features with which the classifiers may perform better. CFsSubsetEval and PrincipalComponentsEval evaluators reduced the features to 6 and 7 respectively. While, SVMAttributeEval ranked the 9 features by using the SVM classifier, then we selected the 7 optimal features. Table 4 shows the selected and ranked feature subsets which will be fed to the classifiers.

Table 4. Features subsets generated by evaluation methods.

Automatic Feature Evaluator	Features Subset
CFsSubsetEval	PWN, NWN, NgWN, PCP, NCP, NWP
PrincipalComponentsEval	PWN, NWN, NgWN, PCP, NCP,CWA, NWP
SVMAttributeEval	PWN, NWN, NgWN, PCP, NCP, RL, NWP

Table 5 shows the accuracy of each classifier based on the outcomes of the used feature evaluator. As noticed, only the SVM classifiers have shown a slight improvement and outperformed the classifiers with the all generated feature subsets, while the performance of other classifiers decreased except NB classifier which showed improvement with feature subset created by PrincipalComponentsEval and SVMAttributeEval. The highest accuracy of SVM classifier obtained with feature subsets generated by CFsSubsetEval and PrincipalComponentsEval. Therefore, they have been investigated in the next phase.

Table 5. The accuracy of each classifier with the three feature subsets.

Classifier	CFsSubsetEval (%)	Principal ComponentsEval (%)	SVMAttributeEval (%)
SVM	90.6915	90.6915	90.5585
NB	76.3298	81.9149	77.6596
RandomForest	90.6915	90.2926	88.1649
K-NN (K=9)	89.8936	89.8936	89.8936

Finally, important to realize that the SVM achieved the highest accuracy through all feature subsets using a linear classifier and without tuning the parameters such as cost parameter and the degree of kernels. Thus, new experiment was carried out to evaluate the SVM classifier through feature subsets with other kernel types and with tuning the different parameters. Kernel types include linear, polynomial, Radial Basis Function (RBF), and sigmoid were applied to the dataset, and the parameters of cost and the degree of the kernel were tuned in order select the model with the highest accuracy. The results of this experiment were summarized in Table 6, and it showed that the highest accuracy obtained with the feature subset generated by PrincipalComponentsEval, and the SVM kernel type was Polynomial when its degree is 2 and its cost is 1.

Table 6. The results of SVM after tuning different parameters.

SVM/Kernels	Accuracy (%)	Precision (%)	Recall (%)	F-Measure (%)
Radial Basis Function (RBF)	89.7606	90	89.8	89.8
Linear	90.6915	90.7	90.7	90.6
Polynomial	92.2872	92.3	92.3	92.3
Sigmoid	76.8617	77.5	76.9	77.1

6. Conclusions and Future Work

In this paper, we presented an approach for sentiment analysis based on sentiment lexicon and machine learning algorithm. Our dataset consists of 2500 reviews written in the Arabic language with the existence of Jordanian dialect. We manually collected the dataset from JEERAN/Jordan website and then pre-processed to be processed by the classifier. We used three automatic feature evaluation methods (CFsSubsetEval, PrincipalComponentsEval, and SVMAttributeEval) embedded in Weka to find the best features subset, to improve the performance of the classification model. In this work, we used four classifiers (NB, SVM, Random Forest, and K-NN) to determine the polarity, and then a comparison of their performances was carried out. The SVM classifier obtained the highest accuracy with features subset recommended by PrincipalComponentsEval.

In the future, the work could be extended to extract and evaluate new features for dialectal Arabic sentiment analysis. We also intend to investigate the potential role that the objective words in the reviews can play. We also plan to investigate the effectiveness of using different methods of feature selection with different vector representations.

References

[1] Abdul-Mageed M., Kübler S., and Diab M., "SAMAR: A System for Subjectivity and Sentiment Analysis of Arabic Social Media," *Computer Speech and Language*, vol. 28, no. 1, pp. 20-37, 2014.

[2] Abdulla N., Ahmed N., Shehab M., and Al-Ayyob M., "Arabic Sentiment Analysis: Lexicon-Based and Corpus-Based," *in Proceedings of IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies*, Amman, 2013.

[3] Aha D., Kibler D., and Albert M., "Instance-Based Learning Algorithms," *Machine Learning*, vol. 6, no. 1, pp. 37-66, 1991.

[4] Al-Subaihin A. and Al-Khalifa H., "A System for Sentiment Analysis of Colloquial Arabic Using Human Computation," *The Scientific World Journal*, vol. 2014, 2014.

[5] Alhumoud S., Altuwaijri M., Albuhairi T., and Alohaideb W., "Survey on Arabic Sentiment Analysis in Twitter," *International Science Index*, vol. 9, no. 1, pp. 364-368, 2015.

[6] Azmi A. and Alzanin S., "Aara'-A System for Mining The Polarity Of Saudi Public Opinion Through E-Newspaper Comments," *Journal of Information Science*, vol. 40, no. 3, pp. 398-410, 2014.

[7] Baly R., El-Khoury G., Moukalled R., Aoun R., Hajj H., Shaban K., and El-Hajj W., "Comparative Evaluation of Sentiment Analysis Methods Across Arabic Dialects," *Procedia Computer Science*, vol. 117, pp. 266-273, 2017.

[8] Breiman L., "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5-32, 2001.

[9] Cambria E., Schuller B., Xai Y., and Havasi C., "New Avenues in Opinion Mining and Sentiment Analysis," *IEEE Intelligent Systems*, vol. 28, no. 2, pp. 15-21, 2013.

[10] Chang C. and Lin C., "LIBSVM: A Library for Support Vector Machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, 2011.

[11] Cleveland R., "A Classification for the Arabic Dialects of Jordan," *Bulletin of the American Schools of Oriental Research*, vol. 171, pp. 56-63, 1963.

[12] Cortes C. and Vapnik V., "Support-Vector Networks," *Machine Learning*, vol. 20, no. 3, pp. 273-297, 1995.

[13] Duwairi R., Marji R., Sha'ban N., and Rushaidat S., "Sentiment Analysis in Arabic Tweets," *in Proceedings of 5th International Conference on Information and Communication Systems*, Irbid, 2014.

[14] Duwairi R., "Sentiment Analysis for Dialectical Arabic," *in Proceedings of 6th International Conference on Information and Communication Systems*, Amman, pp. 166-170, 2015.

[15] El-Beltagy S. and Ali A., "Open Issues in the Sentiment Analysis of Arabic Social Media: A Case Study," *in Proceedings of 9th International Conference on Innovations in Information Technology*, Abu Dhabi, 2013.

[16] ElSahar H. and El-Beltagy S., "A Fully Automated Approach for Arabic Slang Lexicon Extraction from Microblogs," *in Proceedings of Computational Linguistics and Intelligent Text Processing*, Berlin, pp. 79-91, 2014.

[17] Farghaly A. and Shaalan K., "Arabic Natural Language Processing: Challenges and Solutions," *ACM Transactions on Asian Language Information Processing*, vol. 8, no. 4, 2009.

[18] Frank E., Hall M., Holmes G., Kirkby R., Pfahringer B., Witten L., and Trigg L., "Weka-A Machine Learning Workbench for Data Mining," *in Proceedings of Data Mining and Knowledge Discovery Handbook*, Boston, pp. 1269-1277, 2009.

[19] Guyon I., Weston J., Barnhill S., and Vapnik V., "Gene Selection for Cancer Classification Using Support Vector Machines," *Machine Learning*, vol. 46, no. 1-3, pp. 389-422, 2002.

[20] Guyon I. and Elisseeff A., "An Introduction to Variable and Feature Selection," *Journal of Machine Learning Research*, pp. 1157-1182, 2003.

[21] Hall M., Correlation-Based Feature Subset Selection For Machine Learning, Theses, University of Waikato, 1999.

[22] Hetzron R., *The Semitic Languages*, Routledge, 2013.

[23] Ibrahim H., Abdou S., and Gheith M., "Sentiment Analysis for Modern Standard Arabic And Colloquial," *International Journal on Natural Language Computing*, vol. 4, no. 2, pp. 95-109, 2015.

[24] John G. and Langley P., "Estimating Continuous Distributions in Bayesian Classifiers," *in Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence*, Montréal, pp. 338-345, 1995.

[25] Jolliffe I., *Principal Component Analysis*, Springer Science and Business Media, 2002.

[26] Korayem M., Crandall D., and Abdul-Mageed M., "Subjectivity and Sentiment Analysis of Arabic: A Survey," *in Proceedings of Advanced Machine Learning Technologies and Applications*, Berlin, pp. 128-139, 2012.

[27] Liu B., *Sentiment Analysis and Subjectivity*, Handbook of Natural Language Processing, 2010.

[28] Liu B., *Sentiment Analysis and Opinion Mining*, Morgan and Claypool Publishers, 2012.

[29] Maynard D., Bontcheva K., and Rout D., "Challenges in Developing Opinion Mining Tools For Social Media," *in Proceedings of the@ NLP canu tag# usergeneratedcontent*, pp. 15-22, 2012.

[30] McLoughlin L., *Colloquial Arabic (Levantine)*, Routledge, 2009.

[31] Miller C., *Arabic in the City: Issues in Dialect Contact and Language Variation*, Routledge, 2007.

[32] 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*, Atlanta, pp. 55-64, 2013.

[33] Omar N. and Albared M., "Ensemble of Classification Algorithms for Subjectivity and Sentiment Analysis of Arabic Customers' Reviews," *International Journal of Advancements in Computing Technology*, vol. 5, no. 14, pp. 77-85, 2013.

[34] Salim J., "Homonymy in Jordanian Colloquial Arabic: A Semantic Investigation," *English Language and Literature Studies*, vol. 3, no. 3, pp. 69-76, 2013.

[35] Shaalan K., Bakr H., and Ziedan I., "Transferring Egyptian Colloquial Dialect into Modern Standard Arabic," *in Proceedings of International Conference on Recent Advances in Natural Language Processing*, Borovets, 2007.

[36] Stokes J. and Gorman A., *Encyclopedia of the Peoples of Africa and the Middle East*, The Safavid and Qajar dynasties, 2010.

[37] Taboada M., Brooke J., Tofiloski M., Voll K., and Stede M., "Lexicon-Based Methods for Sentiment Analysis," *Computational Linguistics*, vol. 37, no. 2, pp. 267-307, 2011.

[38] Wu X., Kumar V., Quinlan J., Ghosh J., Yang Q., Motoda H., McLachlan G., Ng A., Liu B., Yu P., Zhou P., Steinbach M., Hand D., and Steinberg D., "Top 10 Algorithms In Data Mining," *Knowledge and Information Systems*, vol. 14, no. 1, pp. 1-37, 2008.

Omar Al-harbi is an assistant professor at the Department of Computer & Information at Jazan Community College, Jazan University, Saudi Arabia Kingdom. He Obtained his PhD in Computer Science with specialization in Artificial Intelligence from Islamic Science University of Malaysia (USIM) in 2013. He previously obtained his Master degree in Information Technology from Northern University of Malaysia (University Utara Malaysia UUM) in 2009. Dr. Omar Alharbi also obtained his Bachelor degree in Computer Science from Jerash University, Jordan in 2007. He has over 8 years of teaching experience. His research interests include natural language processing (NLP), word sense disambiguation (WSD), sentiment analysis, and question answering systems (QA).