Journal of Engineering and Applied Sciences 11 (Special Issue 2): 3239-3247, 2016
ISSN: 1816-949X
© Medwell Journals, 2016

Sentiment Analysis in Arabic Social Media Using Association Rule Mining

[1]Ahmed AL-Saffar, [2]Bilal Sabri, [1]Hai Tao, [1]Suryanti Binti Awang,
[1]Mazlina Binti Abdul Majid and [2]Wafaa ALSaiagh
[1]Faculty of Computer System and Software Engineering,
University Malaysia Pahang UMP, Pahang, Malaysia
[2]Data Mining and Optimization Research Group (DMO),
Centre for Artificial Intelligence Technology (CAIT), School of Computer Science,
Faculty of Information Science and Technology,
Universiti Kebangsaan Malaysia (UKM), 43600 Bangi, Malaysia

Abstract: The fast-paced growth in worldwide webs has resulted in the development of sentiment analysis it involves the analysis of comments or web reviews. The sentiment classification of the Arabic social media is an exciting and fascinating area of study. Hence this study brings forth a new method engaging association rules with three Feature Selection (FS) methods in the Sentiment Analysis (SA) of web reviews in the Arabic language. The feature selection methods used are (χ^2), Gini Index (GI) and Information Gain (GI). This study reveals that the use of feature selection methods has enhanced the classifier results. This means that the proposed model shows a better result than the baseline result. Finally, the experimental results show that the Chi-square Feature Selection (FS) produces the best classification technique with a high accuracy of f-measure (86.811).

Key words: Association rule, Arabic sentiment analysis, NLP, machine learning, feature selection method

INTRODUCTION

Currently, many people resort to the online social media for a multitude of purposes. One of them is to give comments or write reviews. The comments or reviews made the online cover a wide range of subjects or interests. These comments or reviews have raised much concern among various parties, especially among users who make purchases online. In addition, the extensive growth of e-commerce has encouraged many users to make online purchases from numerous online markets and stores. When this takes place follow-up activities such as making comments and writing reviews about the service and products become rampant. The act of writing and reading these comments and reviews becomes a platform for users to learn, share and exchange information with one another for optimum choice in making purchases (Saleh *et al*., 2011).

The huge amount of the comments and reviews found online can be positive, negative or neutral in nature. As a result there is a need to classify these comments or reviews. The task of this classification is commonly known as Sentiment classification or sentiment analysis. It is a crucial task for it focuses on classifying the comments or reviews into negative, positive or neutral categories. Various approaches are used for effective sentiment classification. Most of these approaches are confined to machine learning approaches which consist of two parts supervised and unsupervised. Pablos *et al*. (2016) introduced the supervised machine learning approach whereby sentiment or opinion corpora are used in the training of classifiers. Besides, this classification task determines whether the sentiment and opinion present in a specific document, a sentence or phrases (words) as negative, positive or neutral. After the web users write a comment or review about a certain object, it is up to the sentiment classification task to perform the classification based on either weigh upon the entire document or on each individual sentence. Most of the literature studies on Sentiment Analysis (SA) lean on these approaches (Korayem *et al*., 2012; Uysal, 2016). The task of sentiment analysis refers to the field of study that examines the opinions of people about an object, product or services. There are numerous names Sentiment classification or sentiment analysis, namely Opinion Mining (OM) binary classification and classification polarity. Meanwhile, classification polarity aims to identify the polarity of a specified text as negative or positive (Sudhahar, 2015; Saleh *et al*., 2011).

Generally, a majority of the research studies focuses on ways to address the problems faced during the implementation of sentiment classification, in relation to

Corresponding Author: Hai Tao, Faculty of Computer System and Software Engineering, University Malaysia Pahang UMP, Pahang, Malaysia

the reviews and comments. In addition, most of the studies related to sentiment analysis examine texts written in the English language. The main reason lies in the scarcity of resources for language use in the sentiment analysis apart from the English language (Montoyo *et al*., 2012). As well, few research studies examine sentiment classification in Arabic language (Morsy, 2011). This study is based on Arabic sentiment classification models that employ association rule mining and three Feature Selection (FS) methods, namely (χ^2), Information Gain (IG) and Gini Index (GI)). The proposed model evaluates an Opinion Corpus of Arabic (OCA) for (Saleh *et al*., 2011).

The association rule classifier or associative classifier is based on the association rule in a mining task on Data Mining (DM). This classification task is a particular association rule that leads to class label. A frequent itemset is a set of items in which the frequency of these items in a text is more than a threshold value, known as minimum support. From these sets of frequent items, the association rules are derived. These association rules point to the powerful links between the items. They include more contextual information and underlying semantic than an individual word. Moreover, these rules have been developed within the text mining field in diverse aspects (Han *et al*., 2007). Despite the associative classification being well analyzed in text categorization, the application of association rules on sentiment classification issues is still clearly lacking (Man *et al*., 2014).

In addition this study employs Feature Selection (FS) as it is able to reduce word matrix. This reduction marks a crucial phase in the classification method. In sentiment classification, feature reduction or feature selection researches to decrease the dimensionality of the training data sets which can be performed efficiently without affecting the classification results (Ding and Tang, 2013).

Literature review: The use of sentiment analysis has caught much attention among researchers in recent years. As a result, several approaches have been developed for the implementation of sentiment analysis. These approaches are the Machine Learning (ML) approach, lexicon-based approach and a hybrid approach.

Saleh *et al*. (2011) emphasized in examining two machine learning classifers known as Naive Bayes (NB) and Support Vector Machine (SVM) with two weighting terms named Term Frequency (TF) and Term Frequency Inverse Document Frequency (TF.IDF). As well this study used three n-gram representation method. The researchers built their own sentiment corpus by gathering 500 reviews in the Arabic language from multiple websites. They used support vector machine with the tri-gram method without the use of stemmer for the documents in the classifcation. The accuracy rate was reported at 90%. They also claimed that there was no significant effect on the use of TF or TF-ID as a weighting method. This was understandable because both approaches depict the total number of terms over the document. Hence, it is useful to make a comparison between the presence of the terms and terms-frequency method.

Elhawary and Elfeky carried out a sentiment classification on reviews written in the Arabic language on businesses. The findings of their research study assisted in the development of a search engine with numerous mechanisms for skilful annotation of returned pages with sentiment scores. One of the mechanisms comprises a multi-label classifier for the classification of the web pages. This mechanism relies on tagging whereby a tag is assigned to a particular document from a collection of reviews. After they successfully compiled 2000 URLs they performed tagging or labeling manually. It was disclosed that >40% of the URLs were reviewed. The data set was achieved by conducting a web search of keywords and phrases that were part of the review. The final list involves 1500 features that were used in the developing an AdaBoost classifier whereby 80% of the data was controlled item for training purposes and the remaining 20% of the data was set aside for testing purpose.

El-Halees posited a combination of several classifcation methods for the polarity at the document level in the Arabic language. His method was a lexicon-based approach with classifiers Maximum Entropy (ME) and K-Nearest Neighbor (KNN). The result obtained from one classifer served as training data for the next approach. This method involved the use of a simple stemmer for the production of the stem of the Arabic words while the Term Frequency-Inverse Document Frequency (TF.IDF) served as the term weighting. The data f-measure accuracy was used as the evaluation metric. The data domain reported that the f-measure of ranged between 75- 84%. The average of the f-measure was also calculated with 82% for the positive document and 78% for the negative document. As such, the main contention of this study was that the addition of features to the classifer bears no effects to both the accuracy and performance.

Abdul-Mageed *et al*. (2011) designed a Support Vector Machine (SVM) classifier for implementing the subjectivity sentiment analysis on the Arabic genres found in the social media. The data used was gathered

from a major social media. The corpus comprises data in both Modern Standard Arabic language (MSA) and Dialectal Arabic language (DA). The findings of the study revealed the requirement for the development of solutions for every domain and task.

Mourad and Darwish (2013) employed the use of the Naïve Bayes (NB) approach for the extraction of sentiment embedded in a dataset that contained 2300 tweets. The settings of this study included numerous optional features comprise the task in stemming the words found in the tweets. It was reported that the precision of subjectivity detection as in determining whether a tweet is objective or subjective was at 76.6%. Meanwhile, the accuracy of polarity detection as in the choice of whether, if a tweet is negative or positive was at 80.5%.

Ibrahim adopted a feature-based sentence level method for the implementation of sentiment classification in the Arabic language. The Arabic dataset consists of two parts with 1000 tweets in each part. The classification approach was made up of Support Vector Machine (SVM) classifier as well as a lexicon that contained gold-standard sentiment words. The sentiment, word was composed and annotated manually, after which it was extended and the sentiment orientation of the new sentiment words was noted automatically using the synset combination method and free online Arabic dictionaries and lexicons. The results obtained from this approach (SVM with lexicon) revealed that the level of performance was high and the rate of accuracy was precise.

Meanwhile, Duwairi (2015) presented a framework with the ability to analyze sentiments found in tweets written in modern standard Arabic language as well as Arabic dialectical language (Jordanian Dialectical). The tweets were annotated with the help of the crowd sourcing tool. The primary aim of this dialect lexicon framework is to find a match for the dialectal words in modern standard Arabic. The purpose of this study was to ascertain the lexicon value of the dialectal words. The process begins with the starts by the implementation of classification without the use the dialect lexicon. This next step is to use the conversion of the dialectal lexicon into MSA words. The researchers used two machine learning approaches or classifiers for the implementation of sentiment classification. He used Support Vector Machine (SVM) and Naïve Bayes (NB) in his study. The experimental results have demonstrated that the dialectal lexicon indicates some positive effects on the accuracy rate.

The research study by Wang and Wan (2013) applied the association rule mining that was similar to the research reseaech by Hai *et al.* (2011). The recognition of an idea is the ability to recognise a set of fundamental rules inclusive of three practicable ways of extending the set of rules namely the addition of substring rules, dependency rules and constrained topic model rules. The constrained topic model rule had proven to be the most significant as revealed from the finding of the study. The setting of a constraint on the topic model begin with naming one of the feature words, followed by an extension of the topic surrounding that specific word which is also required for the generation of significant clusters. This new way of identifying co-occurrences between the features and other words used in the document complements to the association rule mining technique. The results obtained from this research study were the most desirable with an F1-measure of 75.51% on a dataset of mobile phone reviews written in the Chinese language.

Man *et al.* (2014) adopted a set of association rules for the classification of sentiments found in web reviews. This study examines the datasets containing reviews in of several product types taken from Amazon.com. Each of these reviews has its own features such as product name and a review text. Reviews rated >3 are labeled as positive while those rated <3 are categorized as negative. The reviews that were neither positive nor negative were removed because the polarity is found to be ambiguous. IN fact, the best classification rule set is made to replace the redundant general rule that is of comparatively lower confidence for the implementation of class label prediction procedure. It was recommended that a new metric, Maximum Term Weight (MTW) be included in the assessment of the rules and multiple metric voting schemes with the purpose to overcome the supposed complications arising from inapplicable rules. The final score of a test review depends on the cumulative contributions by the four metrics. The findings obtained from the studies other domain datasets extracted from websites have revealed that the voting strategy on other rule-based processes showed an improvement. In addition, the comparison with the general machine learning approaches indicated that the performance of the recommended technique was better than the strong and tested approaches. In sum, the experimental results have further proved that the strategy applied is successful as it has outperformed the benchmark set by the other approaches.

MATERIALS AND METHODS

This study has suggested a new technique for implementing the analysis sentiment of the Arabic language. This technique is based on a supervised machine learning approach used for adopted the association rule mining. This study outlines the architecture of the Arabic sentiment analysis, together

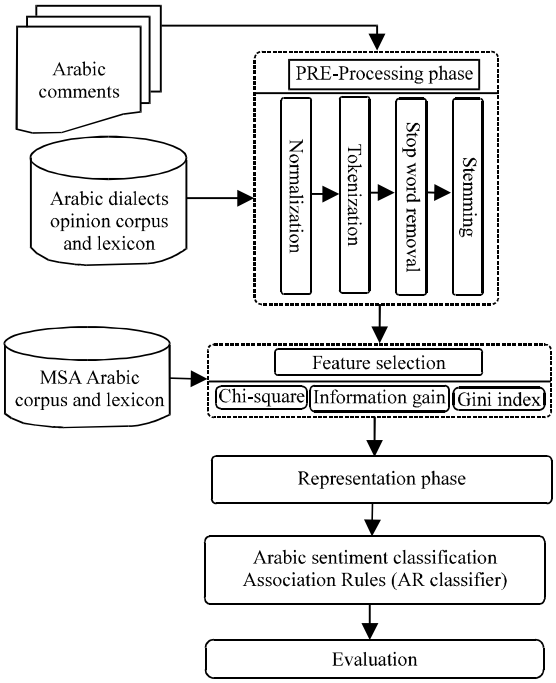

Fig. 1: The architecture of association rule approach for arabic sentiment analysis

with a detailed description of the functionality of each component in the corresponding system. Figure 1 illustrates the architecture of the proposed Arabic sentiment analysis. It also shows the dataset of the Arabic Opinion for Corpus (OCA) which contains 250 positive comments and 250 negative comments. The present research uses also uses the Arabic lexicon as suggested earlier by Al-Moslmi. The reason for the choice of Arabic senti-lexicon is because of its widespread coverage. This senti-lexicon includes 2704 negative, 1176 positive sentiment phrases and words. Moreover, this approach is complete with the following stages:

- Pre-processing stage
- Feature selection stage
- Representation stage
- Sentiment classification stage

The following is a description of each of the stage:

Pre-processing stage: The pre-processing of data is an essential process in text classification, particularly in the sentiment analysis. This is because the high dimension of the Natural Language Process (NLP) of the text makes text data considerably noisy Farzindar and Inkpen (2015). The pre-processing stage of this research study comprises of four phases: normalization, tokenization, stop word elimination and stemming. All of the reviews undergo a pre-processing phase that begins with the normalization procedure. This procedure removes the diacritics, repetitive characters and social media tags. After that is the removal of common stop words. This is necessary as it helps to overcome the problem of the misclassification of a review. Lastly, stemming takes place as it to reduce the term variations to a single representation.

Feature selection stage: Feature engineering is a very significant task in the sentiment analysis domain and it is widely used in text categorization. During this stage, the original documents are changed to feature vectors. As it is a major step in any supervised learning approach for the implementation of sentiment analysis, the researchers have argued that the select on of the right feature set ascertains the whole performance of sentiment classification (Xia *et al*., 2011). As a result, this research examines the strengths and weaknesses of existing features and feature sets. In addition, this research also investigates the effects of feature selection on the performance of Arabic sentiment classification. Likewise Denecke and Deng posit that feature engineering is a very significant stage in the field of sentiment classification and more generally in text categorization. The main aim of the Feature Selection (FS) is to reduce the number of features from the data matrix obtained from the pre-processing stage and removal of unwanted data. This selection process bears some positive effects to enhance the credibility of the text classifier. In addition, the use of FS can mitigate data dimension, improve its accuracy in the removal of noisy features hence expediting the training. Also, Chandrashekar and Sahin (2014) claimed that FS can increase the speed and learning the process of the effectiveness of the various classification tasks. In the feature selection stage, one or more software's are used for processing the training and testing of data to extract descriptive information. For the purpose of sentiment classification, this study has adopted 3 feature selection methods.

Information Gain (IG): The Information Gain (IG) technique is rated highly in the ranking of the most relevant features. Information Gain (IG) measures the relevance of an attribute in relation to class. According to Haddi *et al*. (2013), Information Gain (IG) a commonly known algorithm in Feature Selection (FS), is utilized as a measure for term goodness in the area of Machine Learning (ML). It also estimates the amount of information present in or absent from, a feature. The value obtained from the calculation of IG denotes the rate of accuracy of the classification of a particular class:

$$IG(t) = -\sum_{i=1}^{|c|} p(c_i)\log p(c_i) + p(t)\sum_{i=1}^{|c|} p(c_i|t)\log p(c_i|t) + p(\bar{t})\sum_{i=1}^{|c|} p(c_i|\bar{t})\log p(c_i|\bar{t})$$

Where:

$p(c_i)$ = Probability that class in the occurrence of c_i
$p(t)$ = Probability of the occurrence of the word (t)
$p(t)$ = Probability of the in occurrence of the word t

Chi-square statistic (χ^2): One frequently manipulated feature selection algorithms is the χ^2. This feature selection focuses on the measure of differences between term and group (Yang *et al*., 2009). For the categorization texts, it measures the independence of two random variables; the occurrence of a term, (t) and the occurrence of a class, (c). It is widely used in research studies that focus on the categorization of texts. This is because it is used for the purpose of comparison with other feature algorithms (Tasci and Gungor 2013). The calculation χ^2 value for each term, (t) in category, (c) is based on the following equation:

$$\chi^2(t,c) = \frac{N(AD - CB)^2}{(A + C)(B + D)(A + B)(C + D)}$$

Where N in the equation represents the cumulative count of training reviews in the Arabic language. While A is the count of reviews in the Arabic language of class (c) and it contains term (t), Meanwhile B represents a count of reviews in the Arabic language that do not belong to class (c) but it contains term (t). Subsequently, C is the number of reviews in the Arabic language that do not belong to class (c) and it does not contain term (t). Finally, D is the count of reviews in the Arabic language that does not belong to class (c) and do not contain term (t) (Thabtah *et al*., 2009).

Gini-Index (GI): The feature selection GI is one of the popular methods for the quantification of a feature at the discrimination level. The original application of the form of the (GI) method is for the estimation of the impurity of attributes across classification (Alsaffar and Omar, 2014). An impurity of a smaller value is an indication that the attribute is favorable on the contrary an impurity of a higher value is an indication that the attribute is favorable. The following equation is used to calculate the impurity of this feature:

$$Gini(t) = \sum_{i=1}^{|c|} p(w|c_i)^2 p(w|c_i)^2$$

Based on this method, the highest value, Gini (t) = 1 is obtained if the feature (t) is evident in every document of class, (c_i).

Classification stage: This research study adopts the Association Rule (AR) classifier approach for implementing the sentiment analysis in the Arabic language.

Association Rules (AR): It is common to adopt AR in gathering data, text elements or patterns that co-occur for several times in a dataset. The patterns that are highlighted in the association rules are used for the prediction of data. The association rules employed in this study are based on apriori algorithm. The following describes the steps of the algorithm:

Frequent class terms generation: The common terms used for every category are obtained from the process of generation of frequent terms. Assume a set of terms as T = (t1, t2, t3,..., tm). The term set for the first candidate is obtained directly while the common 1-term set that meets minimum support is retained together with the first candidates to produce the 2-term set. Subsequently, the frequent 2- term set identifies the minimum support for the following candidate. This process of issuing subsequent candidates and frequent m-term sets continuous until no more frequent term set can be produced from the training dataset.

Class association rule generation: The parameters for the application of class association rule are restricted to the rule head and rule body. The application of the rules to the classifier only demonstrates a category label. Generally, the Class Association Rule (CAR) is expressed as $T_j \rightarrow C_i$ where, T_j is a set of frequent terms for the form [t1 and t2 and... and tm] and this set is referred to as rule terms. Meanwhile, C_i being the category of this rule-is identified as the rule head. Each rule must have adequate support and confidence. The rule, CAR is labeled as a strong and frequent rule if it reaches the value of the threshold of minimum confidence (min. confidence) and minimum support (min. support). The support of the rule, $T_j \rightarrow C_i$ is the ratio of reviews for rule terms, T_j which is the cumulative number of reviews in each observed class, C_i. The following equation, support is presented in this ratio:

$$supp(T_j \rightarrow C_i) = \frac{C(T_j \rightarrow C_i)}{N}$$

where by, $C(T_j \rightarrow C_i)$ represents the count of reviews in the dataset where it fits the terms of R, it is related with the category C of R and N is the cumulative count of reviews in class data set. The following equation is for the confidence of rule, $T_j \rightarrow C_i$:

$$CONF(T_j \rightarrow C_i) = \frac{C(T_j \rightarrow C_i)}{C(T_j)}$$

Prediction of new review class: There are so many methods for the adoption of the implementation of classification of reviews based on association rules. This study focuses on the 'ordered decision list (single rule prediction) and 'majority voting (multiple rules prediction) techniques for the classification of reviews. The following describes the steps involved in the majority voting technique. Firstly, it is necessary to go through the rules. Next, identify all the rules that are compatible with the test reviews for classification purpose. After that, check and retained rules are from the same class and then the review is classified to this particular class. Otherwise we will assign this review to a class with most of the retained rules.

Experimental setup: Many experiments have been conducted to assess the proposed model. The initial evaluation is to assess the performance demonstrated by the classification method. The evaluation of the performance of this classification method is based on the Opinion Corpus of Arabic (OCA)

The other evaluation on the performance of the proposed model is the use of the K-fold cross-validation. The main objective of this stage is to fine-tune the parameters in order to opt for the best approach for the classification of sentiment in the Arabic language. For assessing the performance of these classification approaches, the experimental results are categorised into True Positive (TP) which describes sets of reviews that are accurately assigned to the given category. While, False Positive (FP) which describes sets of reviews that are inaccurately assigned to the category and False Negative (FN) which describes sets of reviews that are inaccurate and not assigned to the category. Finally, True Negative (TN) which describes sets of reviews that are accurate but not assigned to the category. Based on the descriptions mentioned the F1 and Macro-F1 measures are adopted for this research study.

$$\text{Prectston} = \frac{\text{TP}}{(\text{TP} + \text{Fp})}$$

$$\text{Recall} = \frac{\text{TP}}{(\text{TP} + \text{FN})}$$

$$\text{F-measure} = \frac{2 \times \text{Recall} \times \text{Precision}}{(\text{Recall} + \text{Precision})}$$

RESULTS AND DISCUSSION

The study of the overall performance of the classifier based on AR for sentiment classification in the Arabic language without the use of FS was implemented on the whole document-term feature space. Table 1 is a summary of the experimental results obtained from the use of association rule classifiers. The experiments were carried out without the use of any feature selection. The best performance is obtained from the use of Association rule classifier.

The effect of the individual feature on the ranking of the performance of the approaches employed by the classifiers is examined at this stage. Table 2-4 show the optimal results obtained when (AR) classifier was applied with three FS (IG, χ^2 and GI) of different feature sizes (100-500). The use of the three FS methods has improved the performance of the association rule compared to the results obtained without the use of the FS methods.

The results tabulated in Table 2 show us the best results with the use of AR classifier with chi-square feature selection of size ranging from (100)-(500). Similarly, the min support used in this study ranges from (10)-(50) and min.Confidence lies in the same range. From Table 2, the highest percentage of accuracy achieved in this experiment is illustrated in row 14 when the size of the applied feature is equal to 500, the min. support is equal to 40 and the min. confidence is equal to 60. The values for precision, recall and f-measure is recorded at 86.800, 86.821 and 86.811 respectively.

Table 3, tabulates the results obtained from the use of association rule classifiers with Information Gain feature selection. The size of the feature used in these experiments ranges from (100)-(500). Meanwhile, the minimum support used in this experiment ranges from 10-50 and minimum confidence lies in the same range as the minimum support. From Table 3, the highest accuracy achieved in this experiment is presented in row 8 when the size of the applied feature size is equal to100, the min. support is equivalent to 20 and the min. is confidence equivalent to 20. The values for precision, recall and f-measure are recorded at 80.800, 80 and 82.787, respectively.

Table 4 tabulates on the experimental results obtained from the use of Association rule classifiers with Gini Index feature selection. The size of the feature used in this experiments ranges from (100)-(500). Meanwhile, the min. the support employed in this experiment ranges from (10)-(50) and the min. Confidence lies in the same range as the minimum support. From Table 4, the highest accuracy achieved in this experiment is presented in row 19 when the size of the applied feature is equal to 500, the min. support is equivalent to 30 and the min. confidence is equal to 60. The values for precision, recall, and f-Measure are recorded at 82.800, 84.146 and 83.468, respectively.

Table 1: Performance (precision, recall, F-measure) of the association rule classifiers

Min. support	Min. confidence	Precision	Recall	F-measure
30	50	74.400	74.406	74.403

Table 2: Performance of the association rule classifiers (precision, recall, f-measure) with chi-square feature selection

Feature-set	Feature type	Min. support	Min.confidence	Precision	Recall	F-measure
500	χ^2	10	30	82.600	82.718	82.659
400	χ^2	10	40	83.800	83.801	83.800
500	χ^2	10	40	85.000	85.005	85.003
400	χ^2	10	50	84.000	84.002	84.001
500	χ^2	10	50	84.200	84.201	84.200
400	χ^2	10	60	82.400	82.402	82.401
500	χ^2	10	60	82.200	82.213	82.206
400	χ^2	20	50	81.600	81.651	81.625
500	χ^2	20	50	82.000	82.074	82.037
200	χ^2	30	50	81.200	81.861	81.529
300	χ^2	30	50	81.400	81.425	81.412
100	χ^2	40	40	81.800	81.862	81.831
400	χ^2	40	60	84.400	84.409	84.404
500	χ^2	40	60	86.800	86.821	86.811
500	χ^2	50	50	81.800	81.862	81.831
100	χ^2	50	60	82.400	82.408	82.404
500	χ^2	50	30	81.600	81.804	81.702
200	χ^2	60	60	82.400	82.419	82.409
400	χ^2	60	60	82.600	82.626	82.613
500	χ^2	60	60	82.600	82.626	82.613

Table 3: Performance of AR classifiers through (precision, recall and F-measure) with Information Gain feature selection

Feature-set	Feature type	Min. support	Min.confidence	Precision	Recall	F-measure
500	IG	10	30	79.600	79.600	79.600
400	IG	10	40	81.200	81.208	81.204
500	IG	10	40	82.400	82.533	82.467
100	IG	10	50	79.800	79.804	79.802
300	IG	10	50	81.400	81.425	81.412
100	IG	10	60	78.800	82.083	80.408
500	IG	10	60	80.000	80.000	80.000
100	IG	20	20	80.800	84.874	82.787
300	IG	20	40	81.600	82.591	82.093
500	IG	20	40	81.200	79.608	80.396
100	IG	20	60	79.200	84.615	81.818
300	IG	20	60	78.800	80.738	79.757
200	IG	30	30	85.600	78.102	81.679
400	IG	30	30	81.200	79.608	80.396
100	IG	30	50	84.000	77.206	80.460
300	IG	30	60	80.000	83.333	81.633
400	IG	40	40	77.200	82.833	79.917
100	IG	40	50	81.600	79.377	80.473
300	IG	50	50	79.200	79.839	79.518
500	IG	50	50	81.200	80.237	80.716

Table 4: Performance of the Association rule classifiers (precision, recall, F-measure) with Gini Index feature selection

Feature-set	Feature type	Min. support	Min.confidence	Precision	Recall	F-measure
200	GI	10	20	81.000	81.112	81.056
400	GI	10	20	80.600	80.604	80.602
400	GI	10	30	80.800	80.808	80.804
500	GI	10	30	81.800	81.801	81.800
400	GI	10	40	83.200	83.219	83.210
500	GI	10	40	82.400	82.533	82.467
100	GI	10	60	78.800	82.083	80.408
400	GI	10	60	82.400	82.072	82.236
500	GI	10	60	84.000	82.353	83.168
100	GI	20	20	79.600	81.557	80.567
500	GI	20	20	80.000	83.333	81.633
300	GI	20	40	82.000	80.078	81.028
400	GI	20	40	81.600	81.600	81.600
500	GI	20	40	81.200	82.186	81.690
500	GI	20	50	79.600	81.557	80.567
400	GI	30	30	81.200	79.608	80.396
300	GI	30	60	80.000	83.333	81.633
400	GI	30	60	84.000	81.712	82.840
500	GI	30	60	82.800	84.146	83.468
500	GI	50	50	81.200	80.237	80.716

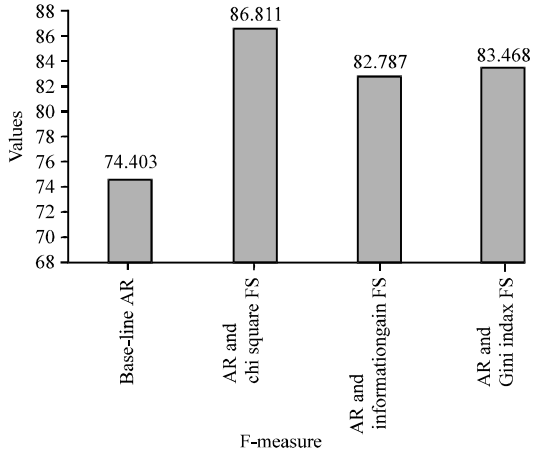

Fig. 2: Comparison of Performances between the classification methods (AR) which are the proposed model with base-line AR classifier

The results listed in Fig. 2 are the best results obtained from the use of AR classifier with chi-square, information Gain, Gini index feature selection and the base-line AR classifier. The results obtained from the use of AR classifier with chi-square feature selection are much better than the results obtained from base-line classifier and AR classifier with the other features selection that has been used in this study for sentiment classification in the Arabic language. These findings confirm that the use of AR classifier approach is the most appropriate method for sentiment classification in the Arabic language.

CONCLUSSION

This study has established a wide comparative study of three FS methods with the use of association rule classification models for the task of sentiment classification in the Arabic language. The leading contribution of this study rested in the examination of the performances of Feature different Selections (FS) and supervised Machine Learning (ML) techniques with special reference to F-measure. The results also show that the best three FS methods have produced improved results as compared to those methods that employed the original classifier. Last but not least, the results reveal that the best performance amongst the FS sentiments is the FS of chi-square while the approach that reached the best performance for sentiment classification in the Arabic language is the AR mining with an f-measure of (86.811).

REFERENCES

Abdul-Mageed, M., M.T. Diab and M. Korayem, 2011. Subjectivity and sentiment analysis of modern standard Arabic. Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Short Papers, Volume 2, June 19-24, 2011, Stroudsburg, PA., pp: 587-591.

Alsaffar, A. and N. Omar, 2014. Study on feature selection and machine learning algorithms for Malay sentiment classification. Proceedings of the International Conference on Information Technology and Multimedia (ICIMU), November 18-20, 2014, IEEE, New York, USA., ISBN:978-1-4799-5423-0, pp: 270-275.

Chandrashekar, G. and F. Sahin, 2014. A survey on feature selection methods. Comput. Electr. Eng., 40: 16-28.

Ding, X. and Y. Tang, 2013. Improved mutual information method for text feature selection. Proceedings of the 8th International Conference on Computer Science & Education (ICCSE), April 26-28, 2013, IEEE, New York, USA., ISBN:978-1-4673-4464-7, pp: 163-166.

Duwairi, R.M., 2015. Sentiment analysis for dialectical Arabic. Proceedings of the 6th International Conference on Information and Communication Systems (ICICS), April 7-9, 2015, IEEE, New York, USA., ISBN:978-1-4799-7349-1, pp: 166-170.

Farzindar, A. and D. Inkpen, 2015. Natural language processing for social media. Synth. Lect. Hum. Lang. Technol., 8: 1-166.

Haddi, E., X. Liu and Y. Shi, 2013. The role of text pre-processing in sentiment analysis. Procedia Comput. Sci., 17: 26-32.

Hai, Z., K. Chang and J.J. Kim, 2011. Implicit Feature Identification Via Co-Occurrence Association Rule Mining. In: Computational Linguistics and Intelligent Text Processing, Gelbukh, A.F. (Ed.). Springer, Berlin, Germany, ISBN:978-3-642-19399-6, pp: 393-404.

Han, J., H. Cheng, D. Xin and X. Yan, 2007. Frequent pattern mining: Current status and future directions. Data Ming Knowl. Discovery, 15: 55-86.

Korayem, M., D. Crandall and M. Abdul-Mageed, 2012. Subjectivity and Sentiment Analysis of Arabic: A Survey. In: Advanced Machine Learning Technologies and Applications, Hassanien, A.E., M.S. Abdel-Badeeh, R. Ramadan and K.T. Hoon (Eds.). Springer, Berlin, Germany, ISBN:978-3-642-35325-3, pp: 128-139.

Man, Y., O. Yuanxin and S. Hao, 2014. Investigating association rules for sentiment classification of Web reviews. J. Intell. Fuzzy Syst., 27: 2055-2065.

Montoyo, A., P.M. Barco and A. Balahur, 2012. Subjectivity and sentiment analysis: An overview of the current state of the area and envisaged developments. Decis. Support Syst., 53: 675-679.

Morsy, S.A., 2011. Recognizing contextual valence shifters in document-level sentiment classification. American University in Cairo, New Cairo, Egypt. http://dar.aucegypt.edu/handle/10526/2351.

Mourad, A. and K. Darwish, 2013. Subjectivity and sentiment analysis of modern standard Arabic and Arabic microblogs. Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, June 14, 2013, Atlanta, Georgia, pp: 55-64.

Pablos, A.G., A.L. Duca, M. Cuadros, M.T. Linaza and A. Marchetti, 2016. Correlating languages and sentiment analysis on the basis of text-based reviews. Proceedings of the Conference on Information and Communication Technologies in Tourism, February 2-5, 2016, Springer, Berlin, Germany, pp: 565-577.

Saleh, M.R., M.T.M. Valdivia, L.A.U. Lopez and J.M.P. Ortega, 2011. OCA: Opinion corpus for Arabic. J. Am. Soc. Inf. Sci. Technol., 62: 2045-2054.

Sudhahar, S., 2015. Automated analysis of narrative text using network analysis in large corpora. Ph.D Thesis, University of Bristol, Bristol, England.

Tasci, S. and T. Gungor, 2013. Comparison of text feature selection policies and using an adaptive framework. Expert Syst. Appl., 40: 4871-4886.

Thabtah, F., M. Eljinini, M. Zamzeer and W. Hadi, 2009. Naive bayesian based on chi square to categorize Arabic data. Proceedings of the 11th IBIMA International Conference on Innovation and Knowledge Management in Twin Track Economies, January 4-6, 2009, IBIMA, Cairo, Egypt, pp: 4-6.

Uysal, A.K., 2016. An improved global feature selection scheme for text classification. Expert Syst. Appl., 43: 82-92.

Wang, W., H. Xu and W. Wan, 2013. Implicit feature identification via hybrid association rule mining. Expert Sys. Appl., 40: 3518-3531.

Xia, R., C. Zong and S. Li, 2011. Ensemble of feature sets and classification algorithms for sentiment classification. Inf. Sci., 181: 1138-1152.

Yang, C.C., Y.C. Wong and C.P. Wei, 2009. Classifying web review opinions for consumer product analysis. Proceedings of the 11th International Conference on Electronic Commerce, August 12-15, 2009, ACM, New York, USA., ISBN:978-1-60558-586-4, pp: 57-63.