

Research Article

Arabic Sentiment Analysis with Optimal Combination of Features Selection and Machine Learning Approaches

Bilal Sabri and Saidah Saad

Data Mining and Optimization Research Group (DMO), Centre for Artificial Intelligence Technology (CAIT), Faculty of Information Science and Technology, School of Computer Science, Universiti Kebangsaan Malaysia (UKM), 43600 Bandar Baru Bangi, Malaysia

Abstract: The main objective of this research study is to design a model that allows for the utilization of a novel technique for the implementation of sentiment analysis in the Arabic language. Sentiment analysis is an interesting task that includes web mining, Natural Language Processing (NLP) and Machine Learning (ML). Most of the research work on sentiment analysis was focused on the texts in the English language. Therefore, the research on sentiment analysis in the Arabic language and other languages are in the infancy stage. This study empirically evaluates three Feature Selection Methods (FSM) (Information Gain (IG), Chi-square (CHI) and Gini Index (GI)) and, three classification approaches (Association Rule (AR) mining and the N-gram model and the Meta-classifier approach) for the implementation of sentiment classification in the Arabic language. A number of related experiments have been carried out on the Opinion Corpus of Arabic (OCA). The results obtained from the experiments were favorable, depending on the algorithms used and the number of selected feature has proven that the use of FS method can increase the performance of sentiment classification in the Arabic language. The results of the experiments reveal that FS method is obtained to develop the classifier performance. Furthermore, the results of the experiment indicated that the use of CHI feature selection has produced the best performance for FS and the performance of meta-classifier a combination approach has outperformed the other approaches for sentiment classification in the Arabic language. In conclusion, this research study has proven that the combination approach (meta-classifier) with the chi-square FS method produces the most accurate classification technique, as high as 90.80%.

Keywords: Feature selection, meta-classifier, machine learning approach, NLP, opinion mining, sentiment analysis

INTRODUCTION

Today, the use of online social media has grown extensively in numerous sectors ranging from a social chat between family members and friends, to doing banking transactions, purchasing fashion wear and to the expression of viewpoints. These online comments or opinions cover a variety of topics in books, movies, electronic products, cars, politics and eateries. This activity has raised the interest of different parties such as customers, companies and government in the analysis and investigation of these opinions. For customers, the rapid growth of e-commerce has encouraged more people to buy from online shops and stores. Consequently, people who began to review the comments about the products had the opportunity to learn from other people's experiences to get a general

idea about these products that can help them in making the best choice (Rushdi-Saleh *et al.*, 2011).

The efficient and accurate mining and identification of useful reviews to meet the needs of both present and potential shoppers have turned into a critical challenge for market-driven product form. As of recent time, data mining and natural language processing in the aspect of correct extraction of people's sentiments from a large quantity of reviews in unstructured text.

Sentiment classification is the main problem with a particular type of text classification that focuses on classifying reviews of overall sentiment polarity into negative or positive categories. There are a variety of techniques and methods for opinion mining or sentiment analysis. Most of the techniques are categorized into two main methods: supervised and

Corresponding Author: Bilal Sabri, Data Mining and Optimization Research Group (DMO), Centre for Artificial Intelligence Technology (CAIT), Faculty of Information Science and Technology, School of Computer Science, Universiti Kebangsaan Malaysia (UKM), 43600 Bandar Baru Bangi, Malaysia

This work is licensed under a Creative Commons Attribution 4.0 International License (URL: <http://creativecommons.org/licenses/by/4.0/>).

unsupervised learning approaches. In the supervised machine learning approach, sentiment corpora are used for the training of classifiers.

There has been an extensive study on ways to overcome the problems faced in the implementation of Sentiment Analysis (SA) based on the comments in the reviews. Most of the works on the analysis of the text are mainly in the English language. The main reason for this occurrence was due to the lack of the resources for use in the sentiment analysis in languages other than English (Montoyo *et al.*, 2012).

In this research study, several sentiment classification techniques have been designed for sentiment analysis in the Arabic language. Also, there is an empirical comparison of three feature selection methods (FSM) (Information Gain (IG), Chi-square (CHI) and Gini Index (GI)). Meanwhile, an investigation of a proposed classification of the FSM has been conducted using three machine learning classifiers (Association Rule (AR), N-gram model and the meta-classifier approach). The evaluation of these approaches is based on an Opinion Corpus of Arabic (OCA) (Rushdi-Saleh *et al.*, 2011).

LITERATURE WORK

The study of sentiment analysis has attracted much attention recently. Several approaches have been developed for the implementation of sentiment analysis. These approaches are the lexicon-based approach, machine learning-based approach and the hybrid approach. Hai *et al.* (2011) proposed the use of Association Rule Mining that recorded the count of co-occurrence of opinion words and explicit features in the Chinese language. The association rule mining is utilized to develop a mapping from the review words to possible features. Likewise, Wang *et al.* (2013) employed the idea of association rule mining that is similar to Hai *et al.* (2011). Both their work involves the identification of a set of basic rules that comprises three feasible ways of extending the set of rules. Man *et al.* (2014) used association rules for the classification of sentiments of web reviews. Their study examines the use of the multi-domain datasets that contain reviews of several product types (fields) taken from Amazon.com. Alsaffar and Omar (2015) applied a model of Malay sentiment analysis that uses sent-lexicon with machine learning approach. They built a new model that generated a set of features for the training of a k-Nearest Neighbor (k-NN) classifier. Omar *et al.* (2013, 2014) investigated several feature selection methods with different machine learning methods. Meanwhile, Soliman *et al.* (2014) utilized a dataset that comprises 1000 tweets (500 tweets in the positive and 500 tweets in the negative) in the Arabic language. The target was the use of sentence-level sentiment analysis based on tweet restriction of 140 characters as a guide. They also

investigated the effects of the use of dialectical language in their research study. As the corpus utilized was rather limited, the researchers appended the use of some Egyptian Arabic language for accuracy. The Support Vector Machine (SVM) and Naïve Bayes (NB) classifiers were utilized in the classification of polarity. Ibrahim *et al.* (2015) conducted a feature-based sentence level approach for Arabic sentiment analysis. The dataset in the Arabic language was made up of two parts; the first part contains 1000 tweets in the dialectical language while the second part consists of 1000 tweets in MSA language. The Support Vector Machine (SVM) classifier used with a lexicon that has been built with gold-standard sentiment words in the case of this classification. The sentiment word was collected and annotated manually, after which the system expands and detects the sentiment orientation of new sentiment words automatically using synset aggregation approach and free online Arabic lexicons and thesauruses.

MATERIALS AND METHODS

This study proposes a supervised machine learning approach for sentiment analysis in the Arabic language. This technique used a meta-classifier combination approach which combines two classifiers named (association rules and n-grams). This section provides an overview of the architecture of the proposed technique in the Arabic sentiment analysis with a description of the functionality of each component in the Arabic sentiment analysis system. Figure 1 presents the architecture of the proposed Arabic sentiment analysis. As well, this technique used Arabic Opinion for Corpus (OCA) dataset, it contains 250 positive comments and 250 negative comments. As well, this approach involves the following phases:

- Data pre-processing phase.
- Feature selection phase.
- Sentiment Classification phase.
- Meta-classification phase.

Pre-processing: Data pre-processing comprises four steps: normalization, tokenization, stop word removal and stemming. All of the reviews involve a pre-processing stage. In the normalization process, diacritics, repeated characters and social media tags are removed. After that, we remove certain stop words that are common in all of the reviews to avoid misclassifying the reviews (Eickhoff, 2015).

Feature selection method: Feature selection or feature engineering is an extremely important task in the domain of sentiment analysis and more generally in text categorization. Converting original documents to

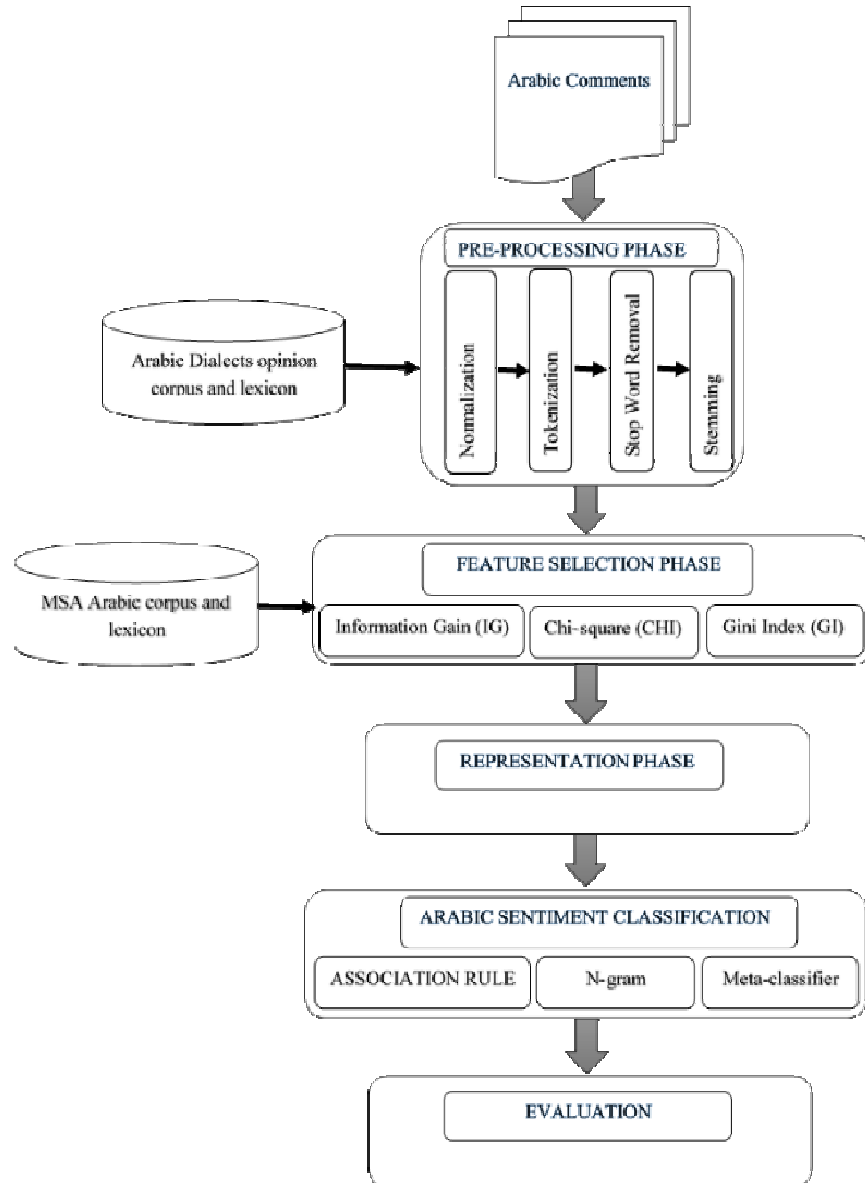


Fig. 1: The architecture of meta-classification approach for Arabic sentiment analysis

feature vectors is the main step in any supervised learning approach to sentiment analysis. Thus, work along this line argues that selecting the right feature set determines the overall performance of sentiment classification (Xia *et al.*, 2011). Consequently, this study investigates the strengths and weaknesses of existing features and feature sets. In addition, this study also studies their effect on the performance of the Arab Sentiment classification. Feature engineering is a very important task in the domain of sentiment analysis and more generally in text categorization. The purpose of the Feature Selection Method (FSM) is to decrease the number of features from data matrix that obtained from pre-processing phase and to remove irrelevant, redundant, or noisy data. The selection process brings positive effects, improving the scalability, efficiency

and accuracy of the text classifier. Additionally, the use of FSM can reduce the dimension of the data, improve accuracy by the removal of noisy features and speed up the training. Also, FSM can improve the speed and learn the effectiveness of classification tasks. In the feature selection phase, one or more types of software are used to process the training and test data for the extraction of descriptive information. This study used 3 feature selection methods for sentiment classification.

Information Gain (IG): The Information Gain (IG) method is employed in the ranking of the most relevant features. Information Gain (IG) measures the relevance of an attribute with respect to a class. Information gain (IG) is a very popular algorithm in Feature Selection (FS). It is utilized as a measure for term goodness in the

area of Machine Learning (ML) (Haddi *et al.*, 2013). IG measures the amount of information that is present in or absent from a feature. The value obtained in the calculation of IG is useful in deciding the correct classification on any class.

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

where $p(c_i)$ the probability that class is, c_i occurs; $p(t)$ denotes the probability that the word, t occurs; and $p(\bar{t})$ is the probability that the word, \bar{t} does not occur.

Chi-square statistic (CHI): One frequently used feature selection algorithms is the CHI. This feature selection focuses on the measure of differences between term and group (Yang *et al.*, 2009). For the tasks of categorizing 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 adopted in studies focused on categorizing texts because it performs well in comparison with other feature algorithms (Taşcı and Güngör, 2013). The CHI. value for each term, t in category, c can be calculated by using the following equations:

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

where, N in the equation represents the cumulative count of Arabic training reviews while A is the count of Arabic reviews in class c and contains term t and B represents a count of Arabic reviews that do not belong to class c but contains term t . Meanwhile, C is the number of Arabic reviews that do not belong to class c and do not contain term t and D is the count of Arabic reviews that do not belong to class c and do not contain term t (Thabtah *et al.*, 2010).

Gini-Index (GI): The Gini-Index is one of the frequently used techniques in the quantification of the discrimination level of a feature. The original form of the Gini Index (GI) algorithm was used in measuring the impurity of attributes across classification (Alsaffar and Omar, 2014). A smaller impurity denotes the best attribute and vice versa. The following equation used to calculate this feature as:

$$Gini(t) = \sum_{i=1}^{|c|} p(w|c_i)^2 p(w|c_i)^2 \quad (3)$$

From the formula, the highest value, $Gini(t) = 1$ will be obtained if feature t appears in every document of class, c_i .

Classification methods: This research study employs three classifier methods namely Association Rules (AR), N-gram model and Meta-classification (stacking) for the implementation of sentiment analysis in the Arabic language. The meta-classifier method is used to combine two classifier AR and N-gram model. This combination method is effective as it is simple to use and it provides accurate results (Franco-Salvador *et al.*, 2015). The following is a description of the classifiers methods.

Association rules: It is common to use association rules in discovering data, text elements or patterns that co-occur for many times within a dataset. The patterns shown in association rules can be used to make predictions of the data. This research study employs the use association rules (Apriori algorithm). The following describes the steps involved in the algorithm:

Frequent class terms generation: The common terms used in every category are obtained during the process of the generation of frequent terms. It is assumed that $T = [t_1, t_2, t_3... t_m]$ is a set of terms. The term set for first candidates are obtained directly and the common 1-term sets that achieve minimum support are retained produce a candidate for the 2-term set. In the next step the frequent 2-term set is identified the following minimum support. The process of producing subsequent candidates and frequent m -term sets continue 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 is limited to the rule head and rule body. The rules which are applied to the classifier only show a category label. Generally, the class association rule (CAR) is expressed as $Tj \rightarrow Ci$, where Tj is a set of frequent terms for the form $[t_1 \& t_2 \& ... \& t_m]$ and this is referred to as rule terms. Meanwhile, Ci which is the category of this rule is known as the rule head. Each rule is accompanied with support and confidence. The rule, CAR is labeled as strong and frequent, if it has achieved the value of the threshold of minimum confidence (min. conf.) and minimum support (min. Sup.). The support of the rule, $Tj \rightarrow Ci$ is the percentage of reviews for rule terms, Tj which is the cumulative number of the reviews in each observed class, Ci . The following presents, the calculation of the ratio for support:

$$supp(Tj \rightarrow Ci) = \frac{C(Tj \rightarrow Ci)}{N} \quad (4)$$

where, $C(Tj \rightarrow Ci)$ represents the count of reviews in the dataset that matches the terms of R and related to category C of R and N is the cumulative count of reviews in class data set. The following is the equation for the calculation of confidence of the rule, $Tj \rightarrow Ci$:

$$CONF(Tj \rightarrow Ci) = \frac{C(Tj \rightarrow Ci)}{C(Tj)} \quad (5)$$

Prediction of new review class: Many methods are used for the classification of reviews based on AR. This research study focuses on two methods of classification of reviews namely ‘ordered decision list (single rule prediction)’ and ‘majority voting (multiple rules prediction)’. The following describes the steps involved in a majority voting technique. The first step is to make a search for the list of rules. The second step is to identify the whole set of rules that matches the test reviews for classification schemes. The third step evaluates all the retained rules as to whether they belong to the same class when the review is classified in this particular class. Otherwise, the review is assigned to a class with a majority of all the retained rules.

N-gram classifier: An N-gram is a technique with an N-character. It slices a long string of words into N-words. Below is a description of the classification using the N-gram classifier.

Generating N-gram frequency profiles: The first step involves the generation of Profiles. It is a very straight forward process that starts with the reading of the incoming text and counts the occurrences of the entire N-grams. Next, the text is divided into tokens that are only made up of letters and apostrophes. After that, it scans each token, in order to generate all possible N-grams, from N = 1 to N = 7. Subsequently, it records the findings into a table to obtain the match for the N-grams. To increase all N-grams and their counts are sorted into reverse order by the number of occurrences. Lastly, the file obtained is an N-gram frequency profile for a review.

Comparing and ranking N-gram frequency profiles: A similar technique is used to measure the distances between every category of profiles (negative and positive) with the new review. Firstly, it is necessary to calculate the profiles of each class (negative and positive). This is followed by the calculation of the closeness of the profiles of many classes for the new review (unknown reviews). The class with the shortest distance are assigned to the reviews. This research study used the cosine similarity technique to measure the similarities of each class profile, CP_i and the profile of the new review, D_j :

$$Sim_{cosine}(CP_i, D_j) = \frac{\sum_{k=1}^m (W_{ik} \times W_{jk})}{\sqrt{\sum_{k=1}^m W_{ik}^2 \times \sum_{k=1}^m W_{jk}^2}} \quad (6)$$

where $k = (1, 2, 3, \dots, m)$ is the number of N-grams in the class profile, CP_i and the profile of the new review, D_j .

Meta-Classification (Stacking): For combination using a meta-classifier, the outputs for all the class labels of component classifier are viewed as new

features for meta-learning. Among the various kinds of classification models, Naïve Bayes (NB) is used to combine the output of the three classifiers. The stacking combination consists of two phases. In the first phase, a set of base-level classifiers is generated. In the second phase, a meta-level classifier is learned that combines the outputs of the base-level classifiers. When using a meta-classifier for combination, the outputs of all the labels of the class of the participating classifiers used as features for meta-learning. In this case, to combine the output of the tow classifiers association rule classifier and N-gram classifier, can use as meta- classifier the Naïve Bayes (NB). The formula of the NB as meta-classifier, given the output of two classifiers O_1, O_2 :

$$P(c_i|O_1, O_2) = \frac{P(c_i).P(c_i|O_1, O_2)}{P(O_1, O_2)} \quad (7)$$

where, $p(c_i|O_1, O_2)$ is the posterior probability of class c_i , given the new output of the two classifiers O_1, O_2 , while $p(c_i)$ is the probability of class.

Experimental setup: Several experiments were conducted to assess the proposed model. One of them was the evaluation of the performance of the classification approach that used Opinion Corpus of Arabic (OCA).

All these algorithms were evaluated through the use of K-fold cross-validation. The stage is necessary as it tunes the parameters for the selection of the best methods for sentiment analysis in the Arabic language. The performance measure of these classification methods, the experimental results were classified into the following: True Positive (TP) is a set of reviews that has been correctly assigned to the category, False Positive (FP) is a set of reviews that is incorrectly assigned to the category, True Negative (TN) is a set of reviews that is correct but it is not assigned to the category. Finally, False Negative (FN) is a set of reviews which is incorrect and it is not assigned to the category. As for this research study, F1 and Macro-F1 measures were being used. The following equation illustrates these matrices:

$$\begin{aligned} Precision &= \frac{TP}{(TP + FP)} \\ Recall &= \frac{TP}{(TP + FN)} \\ F - Measure &= \frac{2 * Recall * Precision}{(Recall + Precision)} \end{aligned} \quad (8)$$

RESULTS AND DISCUSSION

Both classifiers, N-gram model and Association Rule (AR) were initially implemented to the whole document-term feature space to examine the overall

performance of the classifier through the accuracy performance of the sentiment analysis in the Arabic language without used feature selection/reduction methods. Table 1 tabulates the results of the experiments obtained through the use of the Association rule classifiers (AR) and N-gram model. The results of the experiments that were implemented without the use of FS method.

The subsequent sections of this study examine the effects of the individual FS methods on the performance of the classifier. The purpose is mainly to highlight the best results achieved when the (AR) classifier was implemented with Information Gain, Chi-square and Gini index feature selection methods through the use of different features of varying sizes (100 to 500).

This is illustrated in Table 2. The use of the three FSM has enhanced the performance of the Association Rule. This inferred from a comparison of result with and without the use FMS. Table 2 shows that the best in high accuracy performance is shown in line 6 with the application of feature selection, Chi-square of size 500. Also, when the minimum support is equal to 0.40 and the minimum confidence is equal 0.60, the Precision,

Recall and F-Measure were recorded 86.800, 86.821 and 86.811 for respectively.

This is followed by the implementation of the N-Gram classifier to the test set. Table 3 presents the performance of the N-Gram classifier with three FS specifically Gini index, Chi-square and Information Gain in terms of Precision, Recall and F-measure of the sentiment analysis in the Arabic language. Furthermore, the sizes of characters (N) used in the experiments were 3, 4, 5, 6 and 7. The best in high accuracy performance is shown in row 6, with the use of Chi-square FS methods with the size of the feature set of 500. The results of Precision, Recall and F-Measure were recorded at 85.800, 87.6 and 86.7 respectively. Based on analysis of the results, most of the results that achieved high accuracy were obtained when the number of characters (N) is 5 characters or fewer. The reason for this occurrence is most of the words in the Arabic language consists of 5 or fewer characters. This means that when the number of the characters (N) is raised then the accuracy of the performance may be reduced.

Finally, the use of the meta-classifier combination was implemented to the test set. Table 4 tabulates the performance of the application of the combination

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

Model	Precision	Recall	F-measure
N-gram model	73.00	73.083	73.042
Association rule	74.400	74.406	74.403

Table 2: The performances of the use AR approach for sentiment analysis in the Arabic language through the use of three FS with different feature sizes

Feature-set	Supp.	Conf.	Feature. type	Precisions	Recall	F-Measure
500	10	30	CHI	82.600	82.71	82.659
400	10	40	CHI	83.800	83.80	83.800
500	10	40	CHI	85.000	85.00	85.003
400	10	60	CHI	82.400	82.40	82.401
400	40	60	CHI	84.400	84.40	84.404
500	40	60	CHI	86.800	86.82	86.811
500	10	40	IG	82.400	82.53	82.467
100	20	20	IG	80.800	84.87	82.787
100	20	60	IG	79.200	84.61	81.818
400	10	40	GI	83.200	83.21	83.210
500	10	40	GI	82.400	82.53	82.467
400	30	60	GI	84.000	81.71	82.840
500	30	60	GI	82.800	84.14	83.468

Table 3: Performance of N-gram model for sentiment analysis in the Arabic language with the three FS of varying sizes

Feature-set	N-gram	Feature type	Precision	Recall	F-Measure
300	3	CHI	82.400	82.701	82.550
500	3	CHI	83.600	83.967	83.783
400	4	CHI	84.000	85.417	84.702
300	5	CHI	83.600	83.706	83.653
400	5	CHI	84.800	85.181	84.990
500	5	CHI	85.800	87.621	86.701
300	5	IG	75.600	75.602	75.601
400	5	IG	75.600	75.602	75.601
500	5	IG	75.400	75.400	75.400
300	4	GI	75.600	75.602	75.601
200	5	GI	75.800	75.833	75.817
300	5	GI	76.800	76.827	76.814

Table 4: Performance of meta-classifier combination for sentiment analysis in the Arabic language with the three feature selections of varying sizes

Association rules setting		N-gram setting		Meta-classifier		
Feature type	Feature-set	Feature type	Feature-set	Precision	Recall	F-Measure
CHI	300	CHI	300	88.600	89.056	88.827
IG	300	CHI	300	87.800	88.390	88.094
CHI	400	CHI	400	89.800	89.877	89.839
CHI	300	CHI	300	87.400	87.415	87.407
IG	300	CHI	300	87.600	87.600	87.600
CHI	500	CHI	500	90.801	90.810	90.807
GI	500	CHI	500	88.200	88.304	88.252
GI	500	CHI	500	87.800	88.069	87.934
GI	300	IG	300	88.600	88.675	88.637
CHI	400	IG	400	88.200	88.651	88.425
IG	400	GI	400	87.800	88.068	87.934
CHI	500	GI	500	88.200	88.304	88.252

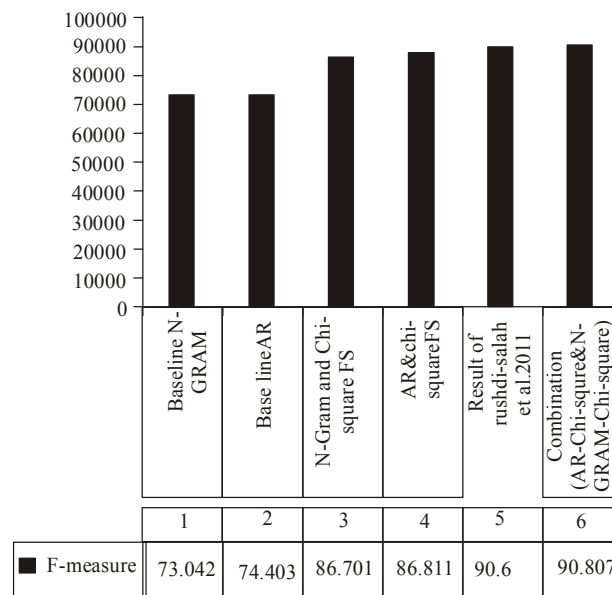


Fig. 2: Comparison of performances of meta-classifier combination with the baseline classifiers, individual classifiers and classifiers used in the previous study

approach (Meta-classifier) with three feature selection specifically Chi-square, Gini index and Information Gain in terms of Precision, Recall and F-measure of the sentiment analysis in the Arabic language. An investigation was conducted on the base classifiers with different settings. Table 4 shows that the employment of the feature selection method had a noticeable effect on the sentiment analysis of Arabic language with the application of combination method (Meta-classifier). The best results were achieved with the use of a meta-classifier approach that is combined (association rule classifier with chi-square feature selection of size 500 and the N-gram classifier with the chi-square feature selection of size 500). The best results for Arabic sentiment analysis from the use of meta-classifier combination was recorded at 90.806, in terms of F-measure.

Generally, the use of Feature Selection (FS) technique has brought positive changes to the performance of all classifiers (Table 1 to 4). Also, the

results have revealed that the meta-classifier combination approach is suitable for the implementation of sentiment analysis tasks in the Arabic language. It is clear that the results obtained from the combination method (Meta-classifier) are better than the results obtained from AR classifier and N-Gram model. This indicates that the proposed technique is a suitable method for the sentiment classification. The results tabulated in Fig. 2 serves as a comparison of the performance of this research study with the previous studies that use the same dataset. The results in Fig. 2 also reveal that the proposed model (meta-classifier) has outperformed that of the baseline classifiers, individual classifiers and classifiers used in the previous study.

CONCLUSION

This study has presented a wide comparative study of three FSM methods and three classification models

for sentiment analysis in the Arabic social media. The principal contribution of this study is the examination of the performance of different Feature Selection Methods (FSM) and machine learning approach in terms of the F-measure. The results also demonstrate that the use of the best three FSM has produced enhanced results in comparison to the results obtained using the original classifier. Among these Fs methods chi-square, FSM yields the best results while the meta-classifier combination shows the best performance with a term f-measure of 90.807 for the sentiment analysis in the Arabic language.

ACKNOWLEDGMNET

This study has been supported by the University research GrantGup-2015-003.

REFERENCES

- Alsaffar, A. and N. Omar, 2014. Study on feature selection and machine learning algorithms for Malay sentiment classification. Proceeding of the IEEE International Conference on Information Technology and Multimedia (ICIMU, 2014). Putrajaya, pp: 270-275.
- Alsaffar, A. and N. Omar, 2015. Integrating a Lexicon based approach and K nearest neighbour for Malay sentiment analysis. *J. Comput. Sci.*, 11(4): 639-644.
- Eickhoff, M., 2015. Enabling Reproducible Sentiment Analysis: A Hybrid Domain-Portable Framework for Sentiment Classification. In: Donnellan, B. *et al.* (Eds.), *New Horizons in Design Science: Broadening the Research Agenda*. Springer International Publishing, Switzerland, pp: 215-229.
- Franco-Salvador, M., F.L. Cruz, J.A. Troyano and P. Rosso, 2015. Cross-domain polarity classification using a knowledge-enhanced meta-classifier. *Knowl-Based Syst.*, 86: 46-56.
- Haddi, E., X. Liu and Y. Shi, 2013. The role of text pre-processing in sentiment analysis. *Proc. Comput. Sci.*, 17: 26-32.
- Hai, Z., K. Chang and J.J. Kim, 2011. Implicit Feature Identification via Co-occurrence Association Rule Mining. In: Gelbukh, A.F. (Ed.), *Computational Linguistics and Intelligent Text Processing*. Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 6608: 393-404.
- Ibrahim, H.S., S.M. Abdou and M. Gheith, 2015. Sentiment analysis for modern standard Arabic and colloquial. *Int. J. Nat. Lang. Comput.*, 4(2): 95-109.
- Man, Y., O. Yuanxin and S. Hao, 2014. Investigating association rules for sentiment classification of web reviews. *J. Intell. Fuzzy Syst. Appl. Eng. Technol.*, 27(4): 2055-2065.
- Montoyo, A., P. Martínez-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(4): 675-679.
- Omar, N., M. Albared, A.Q. Al-Shabi and T. Al-Moslmi, 2013. Ensemble of classification algorithms for subjectivity and sentiment analysis of Arabic customers' reviews. *Int. J. Adv. Comput. Technol.*, 5(14): 77-85.
- Omar, N., M. Albared, T. Al-Moslmi and A. Al-Shabi, 2014. A Comparative Study of Feature Selection and Machine Learning Algorithms for Arabic Sentiment Classification. In: Jaafar, A. *et al.* (Eds.), *Information Retrieval Technology*. Lecture Notes in Computer Science, Springer International Publishing, Switzerland, 8870: 429-443.
- Rushdi-Saleh, M., M.T. Martín-Valdivia, L.A. Ureña-López and J.M. Perea-Ortega, 2011. OCA: Opinion corpus for Arabic. *J. Am. Soc. Inf. Sci. Tec.*, 62(10): 2045-2054.
- Soliman, T.H., M.A. Elmasry, A. Hedar and M.M. Doss, 2014. Sentiment analysis of Arabic slang comments on facebook. *Int. J. Comput. Technol.*, 12(5): 3470-3478.
- Taşcı, S. and T. Güngör, 2013. Comparison of text feature selection policies and using an adaptive framework. *Expert Syst. Appl.*, 40(12): 4871-4886.
- Thabtah, F., Q. Mahmood, L. McCluskey and H. Abdel-Jaber, 2010. A new classification based on association algorithm. *J. Inf. Knowl. Manage.*, 9(1): 55-64.
- Wang, W., H. Xu and W. Wan, 2013. Implicit feature identification via hybrid association rule mining. *Expert Syst. Appl.*, 40(9): 3518-3531.
- Xia, R., C. Zong and S. Li, 2011. Ensemble of feature sets and classification algorithms for sentiment classification. *Inform. Sciences*, 181(6): 1138-1152.
- Yang, C.C., Y.C. Wong and C.P. Wei, 2009. Classifying web review opinions for consumer product analysis. Proceeding of the 11th International Conference on Electronic Commerce (ICEC'09), pp: 57-63.