Do Words with Certain Part of Speech Tags Improve the Performance of Arabic Text Classification?

Abdulmohsen Al-Thubaity
King Abdulaziz City for Science and Technology
Riyadh, KSA
aalthubaity@kacst.edu.sa

Aali Alqarni
King Abdulaziz City for Science and Technology
Riyadh, KSA
aaalqarni@kacst.edu.sa

Ahmad Alnafessah
Department of Computing
Imperial College London
London, United Kingdom
a.alnafessah17@imperial.ac.uk

ABSTRACT
Feature extraction - the process of choosing feature types that can represent and discriminate between dataset topics - is one of the critical steps in text classification and varies with the language of the texts. Different feature types have been proposed for Arabic text classification, ranging from features based on word orthography (single word and character and word N-grams) to features based on linguistic analysis (roots, stems). To the best of our knowledge, little attention has been paid to investigating the performance of Arabic text classification when Part of Speech (POS) tagging information is used to extract features. In this study, we used a corpus comprising 4900 newspaper texts distributed evenly over seven topics to investigate the effect of using POS tag distribution and words that belong to certain POS tags on Arabic text classification, namely nouns, verbs and adjectives. For feature selection, feature representation and text classification we used Chi-squared, Log-Weighted Term Frequency Inverse Document Frequency with Cosine Normalization (LTC) and support vector machine (SVM) respectively. We used four metrics, namely accuracy, precision, recall and F-measure to measure classification performance. Experiment data suggest that the words achieved the best classification performance when the number of features was low; however, the classification performance can be marginally increased when nouns, verbs and adjectives are used as features, given that the number of features is increased.

CCS Concepts
• **Information systems→ Data mining**

Keywords
Arabic text classification; feature selection; POS tagging; classification performance.

1. INTRODUCTION
Automatic text classification is an intelligent way to organize the massive volume of unstructured texts on the Internet and Intranet in private and governmental institutions. The aim of automatic text classification is to assign automatically a given text to one or more predefined classes, based on their content. This text can be a newspaper article, an email, a tweet, a blog or a user opinion on a subject or product, etc. Automatic text classification has several real-life applications, such as sentiment analysis, email classification and spam filtering, or sending news to readers based on the interests that can be learned from their user profiles.

Generally speaking, implementing a text classification system requires five successive main steps: a) collecting and preprocessing a representative dataset; b) defining feature types and extracting them from the dataset; c) selecting representative features and weighing and preparing them for classification algorithm; d) training the classification algorithm and producing the classification model; and e) evaluating the performance of the classification model [9]. Each of the previous steps can easily affect the performance of the text classification system in different ways.

These steps are language-independent except when processing and selecting the feature types which are affected by writing morphological systems of the language. As regards preprocessing, Arabic text classification, for example, may consider the removal of Kashida (ـ) and diacritics and normalising Hamza (أ and إ) and Taa Marbutah (ة). Such preprocessing cannot be applied to any other language. Using word root as a feature type in English text classification, for example, is impossible; however, this is not the case for Arabic. We must, therefore, look for the best preprocessing steps and feature types that may enhance the performance of the classification system according to the used language in the text collection.

The effect on Arabic text classification of using different feature types, such as root, stem, word orthography and word N-grams, has been reported in several studies. Most of these studies showed that word orthography is the best feature type for Arabic text classification [5] [7] [3][1].

In this paper, we will study the effect of using another feature type: Part of Speech (POS) tags. Our survey of Arabic text classification literature has shown a lack of interest in studying the effect of using POS tags on Arabic text classification performance. In this study, we will investigate the effect of using POS tag distribution alone and in conjunction with words, using different scenarios. We will compare the performance results with the results of using word orthography since this is the best-known feature type for Arabic text classification.

The rest of this paper is organized as follows: we summarize the previous efforts of incorporating POS tags information on Arabic text classification in Section II. In Section III, we illustrate the dataset used in our experiments and the experimental setup. The

ICISDM '18, April 9–11, 2018, Lakeland, FL, USA

ACM ISBN 978-1-4503-6354-9/18/04…$15.00
https://doi.org/10.1145/3206098.3206109

results of our experiments and our interpretation of these results are discussed in Section IV. Our conclusions and future work are outlined in Section V.

2. Related Work

Generally speaking, we can divide the features that have been used for Arabic text classification into two types. The first type of features is based on word orthography such as character N-grams [10], single words [9] and word N-grams [2]. The second type is based on linguistic analysis of the single word. These features include roots and stems [5] [7] [8] [1].

In addition to the above-mentioned feature types, two attempts - to the best of our knowledge - have been made to investigate the effect of using POS tagging information on the performance of Arabic text classification. In these two attempts, POS information was used for feature extraction to reduce feature space.

The first attempt is the study conducted by Haralambous [6], in which the authors used the Kalimat corpus as a dataset to investigate the use of stems (using the Stanford Word Segmenter) and roots (as provided by the Kalimat corpus) as features for Arabic text classification. They used Term Frequency Inverse Document Frequency (TFIDF) and sentence parsing information as feature selection methods. For sentence parsing information, they used different strategies that combine the head of the sentence with its neighbour if it is a noun, verb or proper noun. The main conclusions of their study are that stems give a slightly better classification performance than roots, and that parsing information as a feature selection gives, on average, better results.

The second study is the work of Yousif et al [17], in which the authors investigated the use of nouns and adjectives (they did not consider any additional type of POS tags) to enhance the accuracy of Arabic text classification. They used the BBC Arabic Corpus as benchmarking data, word orthography, stem, root and word concept as feature types, and term frequency (TF) as a feature selection method. They compared the performance of text classification with and without using POS tagging and found that incorporating nouns and adjectives marginally enhanced performance when they used word orthography, root (using Root Extractor) and concept relations (using Arabic Word Net). Additionally, their results showed that using nouns and adjectives marginally reduced the classification performance when used with word stem (using the Light stemmer) and root (using the Khoja stemmer).

Studying the effect of using POS tagging information on text classification for other languages has shown contradictory results. Although the use of POS information enhances the accuracy for English text classification [11], the data suggest that incorporating POS tags information does not produces better results than single words for English and Italian TC [15].

Studies by Haralambous et al [6] and Yousif et al [17] show that the use of POS tagging information marginally improves the performance of Arabic text classification; however, several concerns may arise. First, TF and TFIDF are not, usually, the best feature selection methods. Several studies of Arabic text classification studies show that Chi-squared is the best feature selection method [13] [9]. Second, neither of the two studies used all the possible POS tags (nouns, verbs and adjectives).

In this study, we will investigate the effect of using POS tagging information on the performance of Arabic text classification using different scenarios and we will use the best settings for the classification process, as reported in previous studies (see the experimental setting subsection for more details).

3. Materials and Method

3.1 Dataset

Our dataset (SNP 4900) comprised 4900 news articles published in different Saudi newspapers during 2012. Texts were collected manually from the newspapers' websites and classified according to the newspapers' own classification. The dataset text was equally distributed over seven classes (700 texts for each class): economic, cultural, information technology (IT), political, sports, social and general.

To judge the effectiveness of POS tagging on Arabic text classification, we will use two versions of the dataset: a word-segmented version which will be used as a baseline for comparison; and the POS-tagged version of the word-segmented version. We used the Stanford Arabic segmenter [14] for word segmentation. This separates bound morphemes such as possessives, pronouns and discourse connectives (eg, (و، ف، ك، ب، ل، ه، ها، هم). For POS tagging we used the Stanford Arabic POS tagger [16], which uses 33 POS tags. We employed word segmentation for two reasons: first, it improves classification accuracy [3] and second, it improves the POS tagging accuracy for the Stanford Arabic POS tagger. Table 1 illustrates the basic statistics of the dataset before and after segmentation.

Table 1. SNP4900 dataset basic statistics

Classes	Original		Segmented/POS Tagged	
	Total no. of words	No. of unique words	Total no. of words	No. of unique words
Economic	253,307	39,307	313,519	19,629
Cultural	242,070	59,947	307,615	30,270
IT	211,498	41,561	264,671	21,577
Political	188,691	37,223	231,213	19,562
Sports	167,297	31,692	206,247	17,380
Social	158,636	40,112	200,767	21,481
General	110,290	30,911	138,029	19,142
Total	1,331,789	159,944	1,662,061	64,886

3.2 Experimental Setup

In our experiments, we used two versions of the SNP4900 dataset. The first version was the dataset after segmentation, whereas the other was the same dataset after POS tagging. We undertook the following steps to perform our experiments:

1. Preprocessing: Removal of stop words (words that consist primarily of function words in Arabic, such as prepositions, pronouns, conjunctions and possessives), numbers, Latin characters, diacritics and Kashida; normalisation of different forms of Hamza (أ، إ) to (ا) and Taa Marbutah (ة) to (ه).

2. Training and testing size: 10-fold cross validation.

3. Feature type: Four types of features, namely word orthography, words and their POS tags, words that have specified POS tags (nouns, adjectives, verbs and their combinations) and POS tags.

4. Feature selection: We used Chi-squared (CHI) to select and rank the features according to their distinctiveness in each class. CHI has been used in various Arabic

text-classification studies, and its efficiency for feature selection has been demonstrated [9][13]. CHI is calculated as follows:

$$CHI(f_i, c_j) = \frac{N \cdot [P(f_i, c_j) \cdot P(\overline{f}_i, \overline{c}_j) - P(f_i, \overline{c}_j) \cdot P(\overline{f}_i, c_j)]^2}{P(f_i) \cdot P(\overline{f}_i) \cdot P(c_j) \cdot P(\overline{c}_j)} \quad (1)$$

Where:
N is the total number of texts in the training dataset,
c denotes a class,
f is a feature,
$P(c_i)$ is the probability of class c_i, and
$P(t, c_i)$ is the joint probability of class c_i and the occurrence of term t.

5. Feature Representation: We used Log-Weighted Term Frequency Inverse Document Frequency (LTC) to weigh and represent the selected features for the classification algorithm. The weight of feature a_{tj} is given by:

$$a_{tj} = \frac{\log(f(w)+1)\log\left(\frac{Tr}{d(w)}\right)}{\sqrt{\sum_{i=1}^{i=m}\left[\log(f(w)+1)\log\left(\frac{Tr}{d(w)}\right)\right]^2}} \quad (2)$$

Where:
(w) is a word in text t,
Tr is the total number of texts in the dataset,
m is total number of words in the text,
$f(w)$ is the frequency of word w_i in text t, and $d(w)$ is the number of texts t in which word w_i occurs.

6. Classifier training and testing: We used a support vector machine (SVM) for classification. The SVM has proved its efficiency for classification tasks in Arabic text classification [9][12]. We used LibSVM [4], provided by RapidMiner Studio 7.0, to train and test the classification model. We employed 10-fold cross validation to train and text the classifier.

Evaluation: To evaluate the classification performance, we used four measures: Accuracy (Acc), Precision (P), Recall (R) and F-measure (F). The mathematical representations for Acc, P, R and F-measure metrics are illustrated in equations (3)-(6) below:

$$Acc = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

$$P = \frac{TP}{TP+FP} \quad (4)$$

$$R = \frac{TP}{TP+FN} \quad (5)$$

$$F = \frac{2.P.R}{P+R} \quad (6)$$

Where:
TP = true positive
TN = true negative
FP = false positive
FN = false negative

Table 2 lists the experimental parameters.

Table 2. Experimental Parameters

Parameters	Description
Dataset	SNP4900: 7 classes, 4,900 texts. Two versions: not tagged and POS-tagged
Preprocessing	Removing Arabic diacritics, numbers, Latin characters, Kashida and stop-list words; normalising Hamza and Taa Marbutah
Features	Ten types of features: words, words and their POS tags, words with specified POS tags (nouns (N), verbs (V), adjectives (Adj) and their combinations) and POS tags distribution.
Feature selection method	CHI
No. of features Selected	Top-ranked terms (100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100,1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000)
Threshold	Minimum TF = 10
Feature-representation scheme	LTC
Classification algorithm	SVM(LibSVM), 10-fold cross validation (stratified sampling)
Number of experiments	200

4. Results and Discussion

The aim of our experiments is to investigate the effect of incorporating POS tagging information in feature-type selection for Arabic text classification. The results suggest that the use of POS tags distribution only yields a very low classification performance (43.37, 43.08, 43.38 and 0.88 for Acc, P, R and F) due to the low number of features which equal the number of POS tags (33 POS tags). The average Acc, R, P and F for the rest of the feature types are illustrated in Table 3.

Table 3. The average performance results

Feature Type	Acc	P	R	F
W	89.89%	90.17%	89.90%	0.900
W+POS	89.86%	90.15%	89.86%	0.900
N+V+Adj	89.58%	90.10%	89.58%	0.898
N+V	89.27%	89.81%	89.27%	0.895
N+Adj	89.20%	89.83%	89.20%	0.895
N	89.03%	89.57%	89.02%	0.893
V+Adj	83.89%	85.15%	83.90%	0.845
V	75.91%	77.14%	75.95%	0.765
Adj	75.53%	80.06%	75.54%	0.777

As average performance data suggest, using verbs and adjectives only as features yields the lowest classification performance and the combination of the two enhances the classification performance. For the rest of the feature types, the performance results are comparable and no significant difference can be identified. The data suggest that using words as features gives the best classification performance for Acc, R, P and F, followed by words and their POS tags and the combination of noun, verbs and adjectives. It is clear from the data that the classification performance was enhanced dramatically when nouns are incorporated in features.

Figures 1–4, respectively, show the classification performance (Acc, P, R and F), using different top-ranked features for the feature types illustrated in Table 3. The performance data suggest that, as expected, the performance of classification increases as the number of features increases, until the number of features reaches 800, at which point the classification performance becomes almost stable. Furthermore, using words as features gives the best performance results when the number of features is between 100 and 800; the other feature types, namely W+POS and nouns, verbs and adjectives (N+V+Adj), marginally compete when the number of features is between 900 and 2000.

The picture becomes clearer when we consider the best performance results and future type for the top-ranked features and compare these to the performance of the reference feature type (words), as shown in Tables 4-7. The data suggest that using words as features gives better performance results when the number of features is fewer than 800 and when the number of features equals or is greater to 900; other feature types, namely W+POS or N+V+Adj, give marginally better results and using words as features gives the second-best results.

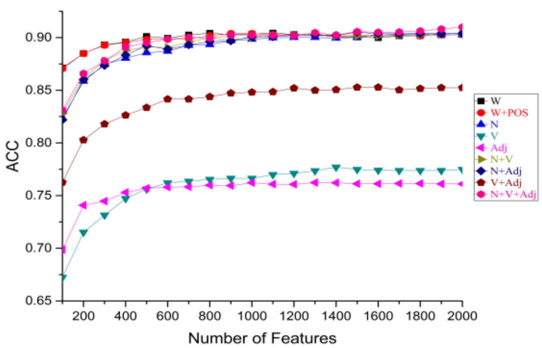

Fig. 1. Classification Accuracy

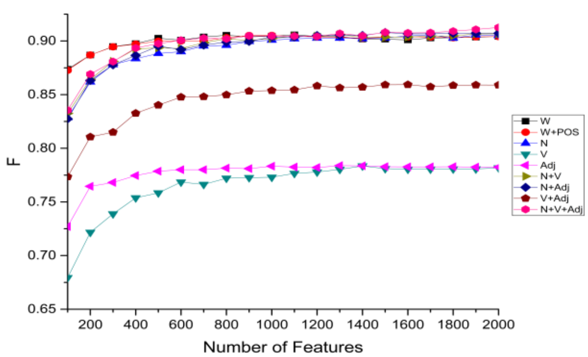

Fig. 2. Classification Precision

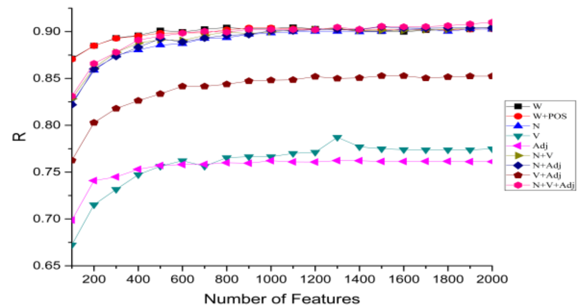

Fig. 3. Classification Recall

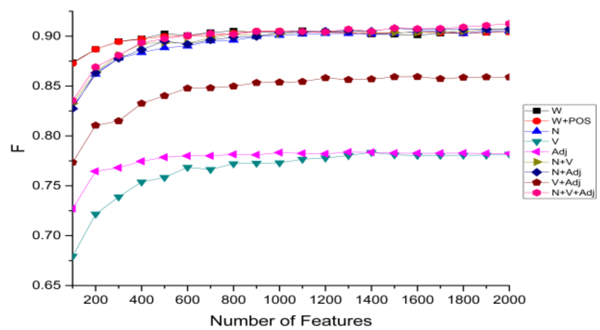

Fig. 4. Classification F-measure

Table 4. The average performance results

# of Features	Reference (W) Accuracy	Best Results Accuracy	Best Results Feature type	Best Results Difference
100	87.12%	87.12%	W	0
200	88.49%	88.49%	W	0
300	89.32%	89.32%	W	0
400	89.59%	89.59%	W	0
500	90.12%	90.12%	W	0
600	89.95%	89.95%	W	0
700	90.26%	90.26%	W	0
800	90.42%	90.42%	W	0
900	90.28%	90.38%	W+POS	0.001
1000	90.34%	90.38%	W+POS	0.0004
1100	90.44%	90.44%	W	0
1200	90.30%	90.30%	W	0
1300	90.28%	90.47%	N+V+Adj	0.0019
1400	90.06%	90.24%	N+V+Adj	0.0018
1500	90.04%	90.57%	N+V+Adj	0.0053
1600	89.98%	90.51%	N+V+Adj	0.0053
1700	90.16%	90.53%	N+V+Adj	0.0037
1800	90.18%	90.63%	N+V+Adj	0.0045
1900	90.24%	90.81%	N+V+Adj	0.0057
2000	90.30%	91.00%	N+V+Adj	0.007

Table 5. The best classification precision and its corresponding feature type

# of Features	Reference (W)	Best Results		
	precision	precision	Feature type	Difference
100	87.61%	87.61%	W	0.000
200	88.90%	88.90%	W	0.000
300	89.64%	89.64%	W	0.000
400	89.91%	89.91%	W	0.000
500	90.40%	90.40%	W	0.000
600	90.20%	90.31%	W+POS	0.001
700	90.49%	90.51%	N+V+Adj	0.000
800	90.63%	90.63%	W	0.000
900	90.50%	90.71%	N+V+Adj	0.002
1000	90.57%	90.70%	N+V+Adj	0.001
1100	90.69%	90.70%	N+V+Adj	0.000
1200	90.55%	90.79%	N+Adj	0.002
1300	90.53%	90.94%	N+V+Adj	0.004
1400	90.32%	90.82%	N+Adj	0.005
1500	90.30%	91.09%	N+Adj	0.008
1600	90.22%	90.99%	N+V+Adj	0.008
1700	90.41%	91.02%	N+V+Adj	0.006
1800	90.43%	91.19%	N+V+Adj	0.008
1900	90.48%	91.35%	N+V+Adj	0.009
2000	90.54%	91.53%	N+V+Adj	0.010

Table 6. The best classification recall and its corresponding feature type

# of Features	**Reference (W)**	**Best Results**		
	recall	**recall**	**Feature type**	**Difference**
100	87.12%	87.12%	W	0.000
200	88.49%	88.49%	W	0.000
300	89.32%	89.32%	W	0.000
400	89.59%	89.59%	W	0.000
500	90.12%	90.12%	W	0.000
600	89.96%	89.96%	W	0.000
700	90.26%	90.26%	W	0.000
800	90.43%	90.43%	W	0.000
900	90.28%	90.39%	W+POS	0.001
1000	90.34%	90.39%	W+POS	0.001
1100	90.45%	90.45%	W	0.000
1200	90.30%	90.30%	W	0.000
1300	90.28%	90.47%	N+V+Adj	0.002
1400	90.06%	90.24%	N+V+Adj	0.002
1500	90.04%	90.57%	N+V+Adj	0.005
1600	89.98%	90.51%	N+V+Adj	0.005
1700	90.16%	90.53%	N+V+Adj	0.004
1800	90.18%	90.63%	N+V+Adj	0.004
1900	90.24%	90.81%	N+V+Adj	0.006
2000	90.30%	91.00%	N+V+Adj	0.007

Table 7. The best classification F-measure and its corresponding feature type

# of Features	**Reference (W)**	**Best Results**		
	F-measure	**F-measure**	**Feature type**	**Difference**
100	0.874	0.874	W	0.000
200	0.887	0.887	W	0.000
300	0.895	0.895	W	0.000
400	0.897	0.897	W	0.000
500	0.903	0.903	W	0.000
600	0.901	0.901	W	0.000
700	0.904	0.904	W	0.000
800	0.905	0.905	W	0.000
900	0.904	0.905	W+POS	0.001
1000	0.905	0.905	W+POS	0.001
1100	0.906	0.906	W	0.000
1200	0.904	0.904	W	0.000
1300	0.904	0.907	N+V+Adj	0.003
1400	0.902	0.905	N+V+Adj	0.003
1500	0.902	0.908	N+V+Adj	0.006
1600	0.901	0.907	N+V+Adj	0.006
1700	0.903	0.908	N+V+Adj	0.005
1800	0.903	0.909	N+V+Adj	0.006
1900	0.904	0.911	N+V+Adj	0.007
2000	0.904	0.913	N+V+Adj	0.008

Tables 8-11 illustrate the best performance results for each type of feature and the corresponding number of features. Data suggest that the combination of N+V+Adj features gives better performance results for all measures when the number of features is 2000 (for accuracy, the number of features is 1900), marginally followed by the combination of nouns and adjectives (N+Adj) when the number of features is 1500. The difference in performance is 0.49%, 0.44%, 0.49% and 0.005 for accuracy, precision, recall and F-measure respectively.

The use of words as features comes third for accuracy and recall and fifth and fourth for precision and F-measure respectively, and the number of features for all best results is 1100. The difference in the performance between feature types N+V+Adj and W is 0.56%, 0.84%, 0.56% and 0.007 for accuracy, precision, recall and F-measure respectively. W+POS features gives comparable performance results with fewer number features whereas the difference in performance between the feature types N+V+Adj and W+POS is 0.62%, 0.89%, 0.61% and 0.008 for accuracy, precision, recall and F-measure respectively, and the number of features equals 900 for accuracy and recall and 1000 for precision and F-measure. The results we obtained are consistent with the results of [9] and [10] in the sense that incorporating POS tagging information can marginally improve Arabic text classification. Therefore, if one is looking for a faster classification system, W+POS features can be used. However, the accuracy, precision, recall and F-measure may be marginally affected. Otherwise, N+V+Adj features may be used to gain better accuracy, precision, recall and F-measure but the classification system may become slower, especially if it is an online system that handles massive amounts of classification requests.

Table 8. The best classification recall and its corresponding feature type

Feature type	Accuracy	# of Features
N+V+Adj	91.00%	1900
N+Adj	90.51%	1500
W	90.44%	1100
N+V	90.42%	2000
W+POS	90.38%	900
N	90.32%	1800
V+Adj	85.28%	1600
V	77.70%	1400
Adj	76.24%	1300

Table 9. The best classification precision and corresponding number of features

Feature type	Precision	# of Features
N+V+Adj	91.53%	2000
N+Adj	91.09%	1500
N+V	90.92%	2000
N	90.83%	1700
W	90.69%	1100
W+POS	90.64%	1000
V+Adj	86.60%	1600
Adj	80.65%	1300
V	79.01%	1400

Table 10. The best classification recall and corresponding number of features

Feature type	Recall	# of Features
N+V+Adj	91.00%	2000
N+Adj	90.51%	1500
W	90.45%	1100
N+V	90.43%	2000
W+POS	90.39%	900
N	90.30%	1900
V+Adj	85.28%	1500
V	78.72%	1300
Adj	76.24%	1300

Table 11. The best classification F-measure and corresponding number of features

Feature type	F-measure	# of Features
N+V+Adj	0.913	2000
N+Adj	0.908	1500
N+V	0.907	2000
W	0.906	1100
N	0.906	1700
W+POS	0.905	1000
V+Adj	0.859	1600
Adj	0.784	1300
V	0.784	1400

5. Conclusion

This study investigated the effect of using POS tagging information on the performance of Arabic text classification, employing different scenarios of combining POS information to extract the features. We compared the use of POS distribution, words and their POS tags, nouns only, verbs only, adjectives only, nouns and verbs, nouns and adjectives, verbs and adjectives, and noun, verbs and adjectives with words without any POS tagging information.

We implemented our experiments on a data set (SNP4900) comprising 4,900 texts equally distributed over seven classes, using Chi-squared for feature selection method, LTC for representation schema, SVM algorithm for classification and Acc, P, R and F-measures as performance metrics. We conducted our experiments for the top-ranked features (100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900 and 2000).

The results of the experiment suggest that using words as features gives a better classification performance for a lower number of features (less than or equal to 800 in our case); however, the classification performance can be marginally improved when nouns, verbs and adjectives are selected as features and the number of features is increased.

6. ACKNOWLEDGEMENT

This research is partly supported by King Abdulaziz City for Science and Technology (KACST) in Kingdom of Saudi Arabia.

7. REFERENCES

[1] Al-Badarneh, A. et al. 2017. The impact of indexing approaches on Arabic text classification. *Journal of Information Science*. 43, 2 (2017), 159–173.

[2] Al-Thubaity, A. et al. 2015. Using Word N-Grams as Features in Arabic Text Classification. *Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing*. R. Lee, ed. Springer International Publishing. 35–43.

[3] Al-Thubaity, A. and Al-Subaie, A. 2015. Effect of word segmentation on Arabic text classification. *2015 International Conference on Asian Language Processing (IALP)*.

[4] Chang, C.-C. and Lin, C.-J. 2011. LIBSVM: A Library for Support Vector Machines. *ACM Trans. Intell. Syst. Technol.* 2, 3 (2011), 27:1--27:27. DOI:https://doi.org/10.1145/1961189.1961199.

[5] Duwairi, R. et al. 2009. Feature Reduction Techniques for Arabic Text Categorization. *J. Am. Soc. Inf. Sci. Technol.* 60, 11 (2009), 2347–2352. DOI:https://doi.org/10.1002/asi.v60:11.

[6] Haralambous, Y. et al. 2014. Arabic language text classification using dependency syntax-based feature selection. *arXiv preprint arXiv:1410.4863*. (2014).

[7] Harrag, F. et al. 2010. Comparing dimension reduction techniques for Arabic text classification using BPNN algorithm. *Integrated Intelligent Computing (ICIIC), 2010 First International Conference on* (2010), 6–11.

[8] Hmeidi, I. et al. 2015. Automatic Arabic text categorization: A comprehensive comparative study. *Journal of Information Science*. 41, 1 (2015), 114–124.

[9] Khorsheed, M.S. and Al-Thubaity, A.O. 2013. Comparative evaluation of text classification techniques using a large diverse Arabic dataset. *Language resources and evaluation*. 47, 2 (2013), 513–538.

[10] Khreisat, L. 2009. A machine learning approach for Arabic text classification using N-gram frequency statistics. *Journal of Informetrics*. 3, 1 (2009), 72–77.

[11] Madsen, R.E. et al. 2004. Part-of-speech enhanced context recognition. *Machine Learning for Signal Processing, 2004. Proceedings of the 2004 14th IEEE Signal Processing Society Workshop* (2004), 635–643.

[12] Moh'd A Mesleh, A. 2007. Chi square feature extraction based svms arabic language text categorization system. *Journal of Computer Science*. 3, 6 (2007), 430–435.

[13] Moh'd Mesleh, A. 2011. Feature sub-set selection metrics for Arabic text classification. *Pattern Recognition Letters*. 32, 14 (2011), 1922–1929.

[14] Monroe, W. et al. 2014. Word Segmentation of Informal Arabic with Domain Adaptation. *ACL (2)* (2014), 206–211.

[15] Moschitti, A. and Basili, R. 2004. Complex linguistic features for text classification: A comprehensive study. *European Conference on Information Retrieval* (2004), 181–196.

[16] Toutanova, K. et al. 2003. Feature-rich part-of-speech tagging with a cyclic dependency network. *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1* (2003), 173–180.

[17] YOUSIF, S.A. et al. Enhancement of Arabic Text Classification Using Semantic Relations with Part of Speech Tagger. *W transactions Advances In Electrical And Computer Engineering*. 195–201.