International Journal of Business, Human and Social Sciences
ISSN: 2517-9411
Vol:9, No:6, 2015

Classification of Political Affiliations by Reduced Number of Features

Vesile Evrim, Aliyu Awwal

***Abstract*—**By the evolvement in technology, the way of expressing opinions switched direction to the digital world. The domain of politics, as one of the hottest topics of opinion mining research, merged together with the behavior analysis for affiliation determination in texts, which constitutes the subject of this paper. This study aims to classify the text in news/blogs either as Republican or Democrat with the minimum number of features. As an initial set, 68 features which 64 were constituted by Linguistic Inquiry and Word Count (LIWC) features were tested against 14 benchmark classification algorithms. In the later experiments, the dimensions of the feature vector reduced based on the 7 feature selection algorithms. The results show that the "Decision Tree", "Rule Induction" and "M5 Rule" classifiers when used with "SVM" and "IGR" feature selection algorithms performed the best up to 82.5% accuracy on a given dataset. Further tests on a single feature and the linguistic based feature sets showed the similar results. The feature "Function", as an aggregate feature of the linguistic category, was found as the most differentiating feature among the 68 features with the accuracy of 81% in classifying articles either as Republican or Democrat.

***Keywords*—**Politics, machine learning, feature selection, LIWC.

I. Introduction

THE explosion in the usage of Internet services moves expressing ideas from conversations to sharing them as text in widely available news/blogs sites and social networking environments. The Information providers on the Web do not necessarily indicate their political affiliations in their writings. Therefore, determining political affiliation of the author of any text recently becomes a challenging issue. Currently, Information Technology researchers are working on extracting political opinions from text documents [1], [26] to analyze the political trends for opinion polling data, to target advertising by sending notices, petitions, donation request or recommending books, clothes or music.

Identifying the political orientation of users is also an interest of the Psychologists. Today, it is observed that people with Liberal view have different interests on topics and different ways of expressing their feelings then the Conservatives [2]. Liberal affiliations are generally characterized by supporting civil rights, democratic elections. On the other hand, Conservative affiliations are generally characterized by preserving the old order, and promoting continuity and stability.

In order to automate the process of determining the political polarity, the research on Machine Learning has gained an importance. The aim in this process is to find out the most differentiating features in political text. The text documents has varieties of features falling into various categories such as linguistics, personality traits, domain related information etc. In this study LIWC, the most commonly used program to find the linguistics and psychological variables in text, was used to select the effective features in determining political orientation [3].

The remainder of this article is structured as follows. In Section II, Background and Related work of political text classification is presented. The corpus construction and Data pre-processing is explained in Section III. Next in Section IV, the experiments including Feature Selection algorithms and Classification algorithms with various numbers of features are explained and discussed. Finally, in Section V, we draw the final conclusions and outline the future work.

II. Background and Related Work

A. Social Media and Politics

The number of opinions aired through the internet has skyrocketed with the increased web publishing. Today, many citizens are involved in concerning policies and applications, perspectives and behavior of politics [4], [5]. The researchers investigate the analysis of opinions and classification of posts aiming to detect new trends in the society according to the understanding, mood, characteristics and behavior of the public [1]. The classification of political articles according to political ideology is also an important component of socio-political studies of political events and its influence on society [1]. Studies that involve opinion mining and classification of political news/blogs usually aim at investigating the possible cause of an event, such as a candidate's failure in elections, by analyzing the tone of statements, sentiment orientation of campaign publications and bias of news reports written about the candidate [6], [25], [26].

B. Politics and Behavior

Environmental conditions are also a factor in defining behavior. Hence, it is reasonable to assume that there is a relationship between politics and personality [7]. Although it has not been long time since the psychological studies of personality extended into the politics, current research shows that personality plays an important role in predicting the results in the domain of politics [8].

Vesile Evrim is with the Computer Engineering Department, European University of Lefke, Mersin 10 Turkey (phone: +90-392-660-2000/2517; fax: +90-392-660-2503; e-mail: vevrim@ eul.edu.tr).

Aliyu Awwal is a M.Sc. student with the Computer Engineering Department, European University of Lefke, Mersin 10 Turkey (fax: +90-392-660-2503; e-mail: aliyu_eul@yahoo.com).

Mairesse et al. investigated the detection of the Big Five aspects of personality in text and audible conversation, utilizing self-rating and observer rating of personality. They tested the classification, regression and ranking frameworks by analyzing the effect of different feature sets on performance accuracy for each framework. The result showed that all statistical algorithms perform much better than the "baseline", although the performance of ranking algorithms performed the best [9].

C. Feature Selection and Classification

One of the main tasks of feature selection algorithms is to select the relevant features that will take a place in the classification process. The first step in feature selection is to test if the features are really relevant or a better model can be obtained by omitting some of the features. Although some of the selected features may not be particularly very relevant, when combined together, there is possibility that selected feature sets would be relevant [10]. The Feature Selection schemes give priority to each feature, according to importance, rank them and then select features with high rankings. This process reduces dimension, time and storage requirements of classifiers while improving the performance and accuracy [11]. Reducing the high dimensional feature space of classifiers is also important to help avoiding the over-fitting problem which is common to highly complex systems [12], [13].

Lee et al. showed that the success of sentence classification is based on only linguistic features [14]. On the other hand, Kotani et al. claimed that text classification marked better results when using both linguistic and learner features [15]. The significance of determining which linguistic features to be applied for experimentation cannot be overemphasized. Many researchers [9], [16] have based their works on LIWC [17] software designed for analysis of texts which have different word classes over an extensive range of dimensions like "positive" or "negative emotions", "self-references", "causal words", as well as seventy other dimensions. Some other researchers [9] used other linguistic systems like MRC [18], which is a "machine-usable" Psycho-linguistic database containing "150,837 words" and 26 "linguistic and psycholinguistic" features.

Classification of political orientation is also studied by different researchers [19]. Pennacchiotti used the Gradient Boosted Decision Tree Learning algorithm on the Twitter user data and found out that topic-based linguistic features are promising in classification of user's political orientation and ethnicity [19], [20]. Monroe identified and evaluated the linguistic differences between Democrats and Republicans in U.S. Senate speeches on a given topic like ''Defense'', ''Taxes'' or ''Abortion'' and illustrate the relative utility of these approaches that base on Bayesian shrinkage and regularization [20].

Therefore in this study, in order to test the political orientation of news/blogs articles, first, the set of features that mostly constitute LIWC categories were used and then the dimension of the vectors was reduced up to a single feature to be tested over many benchmark classification algorithms.

III. Data Collection

The data for this study was collected from online resources with the predetermined affiliations to eliminate the manual classification of the documents. In total, 20 U.S. political blogs and News feeds, as presented in Table XI at Appendix, were RSS fed every 2 hours for 15 days of a period (18/06/2014-02/07/2014), to collect 4000 articles (2000 Liberal, 2000 Conservative). Each of the collected articles was stored in the database with its resource, title, content, and affiliation information and compared against each other to eliminate the possible duplication of articles among the dataset.

A. Data Pre-processing and Feature Vector Construction

One of the aims of this study is to classify the political orientation of the text by using the minimum number of features. Therefore, the LIWC (2007) [17], designed for analysis of texts, which have different word classes over an extensive range of dimensions like "positive" or "negative emotions", "self-references", "causal words", totaling to 64 dimensions were used together with the 4 other features obtained as a result of detailed literature review [9], [16], [21], [22] to constitute the initial feature set of this study.

Each of the 68 dimensions of a feature vector consisted of a set of words as the examples are provided in the third column of Table I. The set of words for the first 64 features were taken from LIWC Library, the words for the 65th and 66th features were extracted from theasearus.com, the set for the 67th feature was extracted from writing.com and the 68th feature is a self-descriptive feature with no predefined set.

All 4000 documents collected from Web were pre-processed through stop-word and punctuation elimination. Although many of the words in LIWC features share common stems, Yarkoni indicates that the relationships between personality and stemmed words could be negatively affected in comparison to un-stemmed words [23]. The reason follows that LIWC features such as "present tense verbs", "past tense verbs", are based on "tenses" and stemming would make it impossible to distinguish between such words [24]. Therefore, in our study, a common pre-processing step, stemming, is not applied and all words are left un-stemmed.

After the document pre-processing, a vector was constructed for each document. The dimensions of a feature vector for each document are made up by calculating the term frequencies (TF) of the words for each feature. Thereafter, the excel sheet with 4000 feature vectors were fed to the RapidMiner Tool for the experimentation as explained in the following section.

IV. Methodology

Identification of the important features for classifying political affiliations is the main goal of our research. To reach this goal several feature selection algorithms were combined with benchmark classification algorithms to find the best

differentiating features of the politic domain. Therefore, the following algorithms were applied on each vector.

- The 10 statistical feature selection algorithms namely, IGR, IG, Correlation, CHI^2, Deviation, Rule, Uncertainty, SVM, Relief, Gini Index and 3 optimized Feature Selection algorithms, Backward Elimination, Forward Selection, Evolutionary were tested over initial set of 68 features. Thereafter, two levels of thresholds were applied on Statistical Feature selection algorithms to reduce the feature dimensions.
- The top selected features of Feature Selection algorithms were tested with 14 classifiers namely, K-NN, Naïve Bayes (NAB), Decision Tree (DET), Rule Induction, Perception(RIN), Neural Networks, ZeroR, M5 Rules (M5R), Gausian, Linear Regression (LIR), Logistic Regression, SVM, Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis, for the performance evaluation.

As a result of the performance evaluation of classification algorithms, the top selected features of the best Feature Selection algorithms were investigated and tested to obtain the best differentiating features for political affiliation classification.

V. Experiments

The task of validating performances for the classification algorithms was done through the X-Validation operator. Experiments on the dataset were run through two set of training/testing ratios:

- 67:33 training/testing ratio with 3 Folds X-Validation
- 80:20 training/testing ratio with 5 Folds X-Validation

In the experiments, Liberal (LIB) class is considered as the positive class (1 binary) and Conservative (CON) class is considered as the Negative class (0 binary). The RapidMiner tool's performance validation operator produces performance results in terms of accuracy, as well as average precision and recall, for positive and negative classes. Since classification is carried out on symmetric datasets, accuracy is considered as the main performance evaluation metric in the study.

A. Full Feature Classification

In this experiment, the 68 features were carried out by 14 classification algorithms of the RapidMiner Tool. The experiments based on both 80:20 and 67:33 training/testing ratios as the accuracy based results are provided in Table II

The results of the "full feature classification" showed that there is not much performance difference between the classifiers' accuracies when they are applied on datasets with 80:20 and 67:33 training/testing ratios. As a result of the experiments, the "M5 Rule", "Rule induction" and "Decision Tree" classifiers outperformed the other classifiers by ~ 82% accuracies, while the "ZeroR", "Gaussian", "Quadratic Discriminant Analysis", "Perception" and "Logistic Regression" models produced the poorest results ~ 50% accuracies which were then removed from the classifier list of the further experiments.

TABLE I
THE COMPLETE FEATURE SET

#	DIMENSION	EXAMPLES
	STANDARD LINGUISTIC DIMENSIONS	
1	Function words	
2	Pronouns	i, them
3	Personal pronouns	her, he
4	1st person singular	i, me
5	1st person plural	we, us
6	2nd person	you
7	Impersonal pronouns	those, it
8	3rd person singular	she, him
9	3rd person plural	they
10	Articles	a, an
11	Adverbs	very, really
12	Verbs	see, run
13	Auxiliary verbs	am, will
14	Past tense	saw, ran
15	Present tense	is, run
16	Future tense	will
17	Prepositions	with
18	Conjunctions	but
19	Swear words	damn
20	Negations	no, not
21	Quantifiers	many
22	Numbers	one
23	SOCIAL PROCESSES	talk, friend
24	Friends	pal, buddy
25	Family	mom, dad
26	Humans	boy, girl
27	AFFECTIVE PROCESSES	happy, ugly
28	Positive Emotions	happy
29	Negative Emotions	sad
30	Anxiety/fear	nervous
31	Anger	kill
32	Sadness	cry
33	CONGNITIVE PROCESSES	know
34	Insight	think
35	Causation	effect
36	Discrepancy	would
37	Tentative	maybe
38	Certainty	never
39	Inhibition	stop
40	Inclusive	with, and
41	Exclusive	except
42	PERCEPTUAL PROCESSES	see, touch
43	Seeing	view
44	Hearing	sound
45	Feeling	hold
46	BIOLOGICAL PROCESSES	eat, pain
47	Body	heart
48	Health	flu
49	Sexuality	love
50	Ingestion	eat
51	Relativity	area, time
52	Motion	move
53	Space	down
54	Time	hour
	PERSONAL CONCERNS	
55	Work	job
56	Achievement	win
57	Leisure	music
58	Home	lawn
59	Money	owe
60	Religion	pray
61	Death	bury
	SPOKEN CATEGORIES	
62	Assent	ok
63	Non-fluencies	uh
64	Fillers	blah
	ADDITIONAL FEATURES TO LIWC	
65	Compliment words	nice, regard,
66	Interjection	hey, oh, ugh
67	Concrete	book, man
68	Words > 6 letters	Complete

TABLE II
THE ACCURACY OF 14 CLASSIFICATION ALGORITHMS WITH 67:33 AND 80:20 TRAINING/TESTING RATIOS

ALGORITHMS	67:33 Train/Test	80:20 Train/Test
K-NN	76.95	77.20
Naïve Bayes	67.75	67.88
Decision Tree	81.32	81.57
Rule Induction	81.95	81.88
Perception	50.23	49.92
Neural Net	76.67	76.75
ZeroR (weka)	49.97	50.00
m5 Rules (weka)	81.65	82.98
Gausian	44.63	45.10
Linear Regression	73.20	73.55
Logistic Regression	57.75	50.00
SVM	77.38	77.72
Linear Discriminant Analysis	73.20	73.55
Quadratic Discriminant Analysis	51.48	51.00

B. Feature Selection Algorithms

In this experiment, 10 Statistical and 3 Optimized Feature Selection algorithms that are available in the RapidMiner tool was used to reduce the dimension of a feature vector, before it was used with classification algorithms. Performance of the feature selection algorithms were tested based on the Average Absolute Deviation (AAD) and the Total Score (TS) values as presented in Table III.

TABLE III
THE AVERAGE ABSOLUTE DEVIATION (AAD) AND TOTAL SCORE (TS) OF FEATURE SELECTION ALGORITHMS OVER 68 FEATURES

ALGORITHMS	AAD	TS
IGR	0.1927	29.12
IG	0.2317	27.16
Correlation	0.1706	36.26
CHI^2	0.1423	10.21
Deviation	0.0768	4.74
Rule	0.1166	8.61
Uncertainty	0.1443	13.08
SVM	0.1352	14.67
Relief	0.0893	9.32
GINI Index	0.2385	28.61

The higher AAD score of a feature selection algorithm is an indication of high differentiation between the features of a selection algorithm which we are interested in this study. Therefore, feature selection algorithms with the lowest AAD scores (Deviation, Rule and Relief) were removed from the set of feature selection algorithms that later be tested with classification algorithms. The top features selected by the remaining feature selection algorithms were further tested in two levels of threshold.

As a result of manual evaluations, the thresholds for the 3 feature selection algorithms, "IGR", "IG", and "GINI index" which have the highest AAD values were determined as 0.3 (Low Threshold) and 0.5 (High Threshold). The threshold values were determined based on the number of features selected by the algorithms. The low threshold value selected to drop the feature size to around 40, while by the high threshold, the number of selected features estimated around 20. The two levels of thresholds for the other feature selection algorithms were calculated proportional to the Total Score value, and the threshold of "IGR", "IG", and "GINI index" algorithms as presented in Table IV.

TABLE IV
TWO LEVELS THRESHOLDS AND THE SELECTED NUMBER OF FEATURES FOR 7 FEATURE SELECTION ALGORITHMS

Algorithms	High Threshold		Low Threshold	
	value	# features	value	# features
IGR	0.50	21	0.30	42
IG	0.50	23	0.30	37
Gini Index	0.50	25	0.30	40
Correlation	0.64	20	0.38	53
CHI^2	0.18	20	0.11	24
Uncertainty	0.23	20	0.14	28
SVM	0.26	20	0.16	39

Optimized feature selection algorithms based the feature selection process to the binary numbers, where 1 indicates the selected features by the algorithms. When 3 optimized feature selection algorithms, "Backward Elimination", "Forward Selection" and "Evolutionary" were tested with 68 features, in an order, 67, 8 and 39 features were selected by the algorithms. Since "Backward Elimination" removed only one, and "Forward Selection" eliminated many of the features by having 8 left, "Evolutionary" algorithm with 39 selected features was selected for later tested by the classification algorithms as explained in the next section.

C. Classification by Selected Features

Classification was carried out on the feature sets selected by the 7 statistical feature selection algorithms as explained in Section V-B. The experiments run on both 67:33 and 80:20 training/testing ratios over a dataset. However, since the classification results for a dataset over the two ratios were similar, with the maximum accuracy difference ~ 1%, for the rest of the experiments, the performance of classifiers were tested with low and high threshold values applied over feature selection algorithms as presented in Table V.

The SVM feature selection algorithm performed the best with 3 classifiers and the second best with the other classifiers when the features were selected by the low threshold, while performing the best on 8 Classifiers with the high threshold. Although the average performance difference of the classifiers among the feature selection algorithms were small (max ~1.93% for low threshold, and max ~0.71% for high threshold), the "Correlation" feature selection algorithm in average performed the worst with classifiers for the features selected by the both low and high threshold values.

As in the results of the full 68 feature classification, the three algorithms namely "M5 Rule", "Rule induction" and "Decision Tree" outperformed the other classifiers on all the feature selection algorithms by varying accuracies 81.5% (±1) for low and high thresholds. The best performance was obtained by "M5 Rule" classifier (82.5%).

The Evolutionary optimized feature selection algorithm was tested with 9 classification algorithms as presented in Table VI. The results showed that the "M5 Rule", "Neural Networks", "Decision Tree" and "Rule Induction" algorithms

had the top 4 accuracies around 80.3% (±0.5). The top classifiers in both statistical and optimized feature selection algorithms were same, except the Neural Network Classifier which had 1% increase in its performance by optimized feature selection algorithm. Although no big difference was obtained between the statistical and optimized algorithms, it was observed that the accuracies of the top classifiers dropped ~ 1% by the Evolutionary optimized feature selection algorithm.

TABLE V
THE ACCURACY OF 9 CLASSIFIERS WITH 7 FEATURE SELECTION ALGORITHMS WITH LOW AND HIGH THRESHOLDS

Algorithms	K-NN	Naïve Bayes	Decision Tree	Rule Induction	Neural Network	m5 Rule	Linear Regression	SVM	LDA	Average
80:20 training/testing with Low Threshold										
IGR	77.48	67.98	81.60	81.68	76.75	82.12	73.32	77.52	73.30	73.95
IG	77.12	67.10	81.50	81.88	76.90	82.02	72.93	65.70	72.80	72.00
GINI INDEX	77.18	67.40	81.10	81.80	74.30	82.48	73.25	74.20	73.18	72.32
CORRELATION	77.60	68.08	81.38	81.65	76.22	81.12	73.87	77.68	73.87	72.02
UNCERTAINITY	76.65	68.52	79.80	81.42	73.52	80.50	71.73	73.28	71.48	72.47
SVM	76.65	68.08	81.45	81.77	79.60	81.72	74.00	77.62	74.08	73.04
CHI SQ	76.70	68.65	79.80	81.32	79.85	81.42	71.80	73.55	71.70	72.94
Average	77.05	67.97	80.95	81.65	76.73	81.63	72.99	74.22	72.92	72.68
80:20 training/testing with High Threshold										
IGR	76.75	66.78	81.50	80.98	78.12	81.40	70.90	74.02	70.80	72.38
IG	76.75	66.55	81.50	80.98	79.47	81.05	70.68	76.62	70.75	72.52
GINI INDEX	76.98	66.72	81.50	80.98	78.82	81.48	71.02	77.40	71.43	73.06
CORRELATION	76.45	68.32	81.50	80.98	79.00	81.95	71.58	71.65	71.78	72.29
UNCERTAINITY	76.32	68.60	79.80	80.78	80.22	81.05	70.55	71.80	70.48	73.01
SVM	77.35	68.92	81.52	80.98	81.00	82.32	73.10	77.18	73.08	73.09
CHI SQ	76.32	68.60	79.80	80.78	80.22	81.05	70.55	71.80	70.48	73.01
Average	76.70	67.78	81.02	80.92	79.55	81.47	71.20	74.35	71.26	72.77

D.Discussion of the First Test Results

As results of the experiments, unexpectedly, a small difference was observed between the classification results of the full 68 features, and the classification results based on the subset of features selected by statistical and optimized feature selection algorithms. Table VII shows the performance differences between the full 68-feature set classifications, and the best results obtained for each classifier over the low and high threshold values of 7 feature selection algorithms. The table also shows the comparison of 68-feature set classifications and classifications applied on evolutionary algorithm.

TABLE VI
THE ACCURACY OF EVOLUTIONARY OPTIMIZED FEATURE SELECTION ALGORITHM OVER 9 CLASSIFIERS

Classifiers	Accuracy
K-NN	76.88
Naïve Bayes	67.52
Decision Tree	80.15
Rule Induction	79.88
Neural Networks	80.65
m5 Rules (weka)	80.80
Linear Regression	73.42
SVM	77.12
Linear Discriminant Analysis	73.58

Overall, we obtained that there is no big difference between the results of classifiers on statistical feature selection algorithms in two levels of thresholds, and the Evolutionary optimized feature selection algorithms. In both categories, algorithms had an accuracy performance around 80%. Similarly, with the exception of the Neural Networks the results showed that, the accuracy of classifiers tested with the 68 features and the features selected by the low and high threshold values varied about 1%. It is also observed that the success of Neural Network classifier increased up to 4.25% as the dimension of the selected features decreased.

TABLE VII
THE COMPARISON OF THE BEST PEROFRMANCES OF CLASSIFIERS ON LOW/HIGH THRESHOLD OF STATISTICAL FEATURE SELECTION ALGORITHMS, OPTIMIZED EVOLUTIONALLY ALGORITHM WITH 68-FEATURE SET CLASSIFICATIONS ON 80:20 TRAINING/TESTING RATIO DATASET

ALGORITHMS	Low	High	Optimized
K-NN	0.40	0.15	-0.32
Naïve Bayes	0.77	1.04	-0.36
Decision Tree	0.03	-0.05	-1.42
Rule Induction	-0.11	-0.90	-2.00
Neural Networks	3.10	4.25	3.90
m5 Rules (weka)	-0.50	-0.66	-2.18
Linear Regression	0.45	-0.45	-0.13
SVM	-0.04	-0.54	-0.60
Linear Discriminant Analysis	0.03	0.53	-0.47

E. Classification over Reduced Feature Set

Since reducing the number of features in a vector from 68 to 20 (High Threshold) did not have noticeable effect on the performances of the classification algorithms, we decided to reduce the dimension of a feature vector even more to find out the most important features affecting the performances of the classifiers.

TABLE VIII
FIRST 15 FEATURES SELECTED BY 7 FEATURE SELECTION ALGORITHMS

N	IGR/IG/GINI	Correlation	Uncertainty	SVM	CHI^2
1	Function	Words>6	Quantifier	Words>6	Quantifier
2	Words>6	Concrete	Pronoun	Hearing	Pronoun
3	Preposition	Article	Words>6	Relativity	Words>6
4	Article	Interjection	Work	2ndpp	Work
5	Verb	Preposition	Hearing	Verb	Concrete
6	Cognitive	Space	Concrete	Cognitive	Insight
7	Relativity	Relativity	Motion	Negation	1stpsp
8	Pronoun	Function	Insight	Certainty	Motion
9	Auxiliary	Compliment	1stpsp	Concrete	Article
10	Conjunction	Inclusive	Article	Number	Home
11	Concrete	Conjunction	Adverb	Positive	Hearing
12	Space	Time	Achieve	Family	Adverb
13	Present	Hearing	Money	1stpsp	Achieve
14	Impersonal	Verb	Home	Pronoun	Inhibition
15	Inclusive	Perceptual	Interjection	Tentative	Money

In order to have a fair comparison between the feature selection algorithms, the top 1^{st}, 5^{th}, 10^{th} and the 15^{th} selected features of the feature selection algorithms, as presented in Table VIII, were tested against 6 Classifiers. The three classifiers namely "KNN", "Neural Networks" and "Linear Regression" were eliminated from the list of classifiers because of their poor performance and high time complexity. Table IX shows the performances of the classifiers on the different feature selection algorithms with various numbers of features.

The "IGR", "IG" and "GINI Index" feature selection algorithms shared the same top 15 features which performed the best with the "Decision Tree" and "Rule Induction" classifiers. Most of the other classifiers gave the best results with "SVM" feature selection algorithm. The three features "function", "quantifier", "words > 6", were ranked as the top features by the 7 feature selection algorithms. Among these three features, the feature "quantifier" performed the worst with all classifiers while the "function" feature performed the best for the top four classifiers.

The "CHI^2" and "Uncertainty" feature selection algorithms shared the same top four features. Therefore, the classification on the top 5 features set of these algorithms revealed information about the relevancy of the 5^{th} features that are "Concrete" and "Hearing". Although the classifiers performance differences for the features "Concrete" and "Hearing" found negligible, the feature "Hearing" had a better effect on all classifiers compared to "Concrete". Therefore, the top 5 features of "Uncertainty" feature selection algorithm which includes "Hearing" feature performed up to 2.2% better than "CHI^2" algorithm when used with Naïve Bayes classifier.

The results showed that the performances of the 64 feature classifications were not more than 3% higher compare to the classification with a single feature for the best four classifiers. On the other hand, when the top 5 features of the feature selection algorithms were tested by the classifiers, the difference with the 64 feature classification results has decreased to ~1% in many cases. Surprisingly, "Naive Bayes" classifier, with the top 5 features of "SVM" feature selection algorithm, had its best performance over the dataset with 3% improvement to 64 feature classification.

The accuracy of the classifiers when tested with the top 10 and the top 15 feature sets did not show any noticeable differences. The performance of "SVM" classifier showed steady increase up to 6% between the top feature, and the top 15 feature sets. Similarly, the "M5 Rule", "SVM" and "Linear Regression" classifiers when used with "SVM" feature selection algorithm performed in an order 0.4%, 0.4%, and 1.2%, less than the performances of the classifiers used 64 features.

TABLE IX
PERFORMANCE OF THE CLASSIFIERS WITH DIFFERENT FEATURE SELECTION ALGORITHMS AND VARIOUS NUMBERS OF FEATURES

#Features	Algorithms	NAB	DET	RIN	M5R	SVM	LIR
1	SVM/Corr	67.27	79.80	79.02	80.25	69.88	68.88
	Uncer/Chi^2	63.63	73.80	54.32	73.40	64.42	66.95
	IGR/IG/Gini	66.45	81.00	81.32	81.45	70.10	68.15
5	SVM	70.15	80.18	80.95	81.40	71.48	71.28
	Correlation	67.10	79.80	80.95	80.93	70.30	68.90
	Uncertainty	69.05	79.80	81.20	81.40	70.75	69.90
	Chi^2	66.82	79.80	81.18	81.08	70.38	69.00
	IGR/IG/Gini	67.00	81.35	81.62	81.95	70.10	69.10
10	SVM	69.15	80.10	80.78	81.40	75.50	72.42
	Correlation	67.02	81.40	81.83	81.62	69.98	69.23
	Uncertainty	68.50	79.80	81.38	81.05	71.98	70.17
	Chi^2	67.00	79.80	81.25	81.32	71.68	69.08
	IGR/IG/Gini	66.85	81.18	81.65	81.32	70.15	69.48
15	SVM	69.10	80.08	81.23	82.40	76.13	72.35
	Correlation	68.48	81.60	81.85	81.15	70.98	70.43
	Uncertainty	68.70	79.80	81.38	80.60	71.85	70.43
	Chi^2	68.58	79.80	81.52	80.38	71.70	70.18
	IGR/IG/Gini	66.78	81.35	81.75	81.50	71.40	70.48

In addition to the accuracy measure, during the feature reduction process, the other performance measures such as precision and recall were also analyzed. Compare to the performances of the classifiers with 68 feature set, "SVM" classifier had as much as 8% decrease in precision and 16% in recall when tested with a small set of features except the outlier top feature of "Uncertainty" algorithm that dropped the recall value of the "SVM" classifier as low as 39%.

As results of experiments it is observed that classification with the reduced set of features, moreover with a single feature, maintains up to 80% accuracy in differentiating the political orientation of the text as Republican or Democrat. The closer look to the selected features indicated that the most of the top selected features corresponds to the linguistic dimensions of the initial feature set. For example, from the top five features selected by the 7 feature selection algorithms, the "IGI", "IG" and "GINI Index" selected 5 as linguistic features, whereas "Correlation", "Uncertainty", "SVM", and "CHI^2" selected 3 as linguistic features.

F. Linguistic Feature Set

By the guidance of the results obtained from the experiments, we made more investigation to see the effects of the linguistic dimensions over political affiliation

classification. The top 20 features of the best performing feature selection algorithms, "SVM" and "IGR", were analyzed and the subset with 10 linguistic features were extracted for further experimentation.

As the best performing top 3 classifiers of the previous experiments, the "Decision Tree", "Rule Induction", "M5 Rule" and the popular "SVM" classifier which has balanced precision and recall values, were selected to be tested over linguistic features. The selected 10 linguistic features were first tested individually and then combined and tested all together over 4 classifiers which performed ~81.5% accuracy as presented in Table X.

TABLE X
THE ACCURACY OF INDIVIDUAL AND SET OF LINGUISTIC FEATURES OVER 4 CLASSIFIERS

FEATURES	DET	RIN	M5R	SVM
10 Lexical	81.63	81.88	81.80	72.68
Words > 6	79.72	78.05	79.12	69.85
Verbs	78.95	60.73	78.85	70.62
Pronouns	79.12	60.60	78.82	70.68
Number	50.00	51.00	67.52	60.55
Preposition	80.02	67.07	79.40	69.98
Function	80.92	81.00	81.50	70.02
Conjunction	78.35	58.62	77.85	67.92
Concrete	77.70	61.60	77.02	69.92
Article	79.20	60.40	79.50	70.35
2'nd Person	50.00	52.05	66.68	58.43

The results of classifications with 10 lexical features showed similarity to the results of classifiers used with the top 10 features of feature selection algorithms by improving ~0.2% for 3 classification algorithms. The single lexical features in general performed worse. As one of the top selected features by feature selection algorithms, the lexical feature "Function" performed ~80% and the feature "words > 6" performed ~79%. The outstanding performance of a single feature "function" is not surprising since the LIWC feature "Function" includes the set of words that are aggregate of the other 21 linguistic features. The experiments also showed that the lexical features "Number" and "2'nd Person" performed the worst by the best accuracies of 67%.

VI. CONCLUSION AND RELATED WORK

In this study, various experiments were conducted to find the best reduced set of features without losing the performance of classifiers. Several feature selection algorithms and feature selection methodologies were used to reduce the dimension of the feature space. The best performance (82.98%) was obtained by the "M5 Rule" classification algorithm with the 68 feature set. The "M5 Rule" algorithm with the features selected by "GINI Index" on low threshold (40 features) performed 82.48% and again the "M5 Rule" over the top 15 features selected by "SVM" feature selection algorithm obtained 82.40% accuracy. Moreover, the classification of a single feature "function" with "M5 Rule" performed 81.45%.

Overall, the "M5 Rule" classifier had the best performance in all the experiments. Majority of the classifiers performed ~80% accuracy over the dataset. The difference between the performances of the "M5 Rule" classifier with the single feature "Function" and the 68 feature set is obtained as 1.48% that is within the error margin of the validation in the RapidMiner tool. Although the foundations were not too strong to say that linguistic features are better determinants of the political orientation, they provide an evidence to investigate in the future research.

APPENDIX

TABLE XI
THE 20 U.S. POLITICAL BLOGS AND NEWS FEEDS

	Liberal Sources
1	http://www.huffingtonpost.com/feeds/verticals/politics/index.xml
2	http://www.dailykos.com/rss/Diary.xml
3	http://feeds.feedburner.com/latimes/news/opinion/commentary?formatxml
4	http://www.thenation.com/blogs/rss/politics
5	http://www.thenation.com/rss/blogs
6	http://feeds.washingtonpost.com/rss/rss_election-2012
7	http://feeds.washingtonpost.com/rss/rss_right-turn
8	http://feeds.feedburner.com/Motherjones/mojoblog?format=xml
9	http://feeds.feedburner.com/motherjones/Politics?format=xml
10	http://thinkprogress.org/election/issue/feed/
	Conservative Resources
11	ttp://z.about.com/6/o/m/usconservatives_p2.xml
12	http://www.washingtontimes.com/rss/headlines/news/politics/
13	http://www.washingtontimes.com/rss/weblogs/inside-politics/
14	ttp://nypost.com/opinion/feed
15	http://feeds.feedburner.com/michellemalkin/posts?format=xml
16	http://feeds.feedburner.com/hotair/main
17	http://spectator.org/feed
18	http://www.theamericanconservative.com/articles/feed
19	http://pamelageller.com/feed
20	http://www.theamericanconservative.com/articles/feed

REFERENCES

[1] M. Kaschesky, S. Pavel, and B. Guillaume, "Opinion Mining in Social Media: Modeling, Simulating, andVisualizing Political Opinion Formation in the Web," 2012.
[2] Y. Inbar and L. Joris, "Perspectives on Psychological Science," 2012.
[3] J. W. Pennebaker, R. E. Boot, and M. E. Francis, "Linguistic inquiry and word count: LIWC2007 - Operator's manual," Austin, TX, 2007.
[4] R. Inglehart and C. Welzel, "Modernization, Cultural Changeand Democracy", Cambridge UK, 2005.
[5] M. Griffiths, "E-citizens: Blogging as democratic practice", 2004.
[6] Y. Fang, L. Si, N. Somasundaram, and Z. Yu, "Mining Contrastive Opinions on Political Texts using Cross-Perspective Topic Model," in ACM, 2012, pp. 1-15.
[7] D. W. Van,"Shockmd: a neurostimulating blog". (Online). http://www.shockmd.com/2009/12/16/personality-traits-and-political-attitude/ , 2009.
[8] S. Alan Gerber, A. Gregory Huber, David Doherty, and Conor M. Dowling, "Personality and Political Attitudes: Relationships across Issue Domains and Political Context", vol. 104(1), 2010, pp. 111-133.
[9] F. Mairesse, M. A. Walker, M. R. Mehl, and R. K. Moore, "Using Linguistic Cues for the Automatic Recognition of Personality in conversation and text," Journal of Artificial Intelligence Research, 2007 pp. 457-500.
[10] S. Marina and P.W.D. Robert, "Combining feature subsets in feature selection," Multiple classifier systems, 2005 pp. 165-175.
[11] E. Ozbilen, "improving text categorization performance by combining feature selection methods," Istanbul, 2008.
[12] G. Forman, "An extensive empirical study of feature selection metrics for text classification," Journal of Machine Learning Research, 2003 vol. 3, pp. 1289–1305.

[13] E. R. Dougherty, J. Hua, and C. Sima, "Performance of Feature Selection Methods ," Current Genomics, vol. 10(6), 2009, pp. 365–374.
[14] J. Lee, M. Zhou, and X. Liu, "Detection of non-native sentences using machine-translated training data," in Proceedings of the 2007 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics, 2007, pp. 93-96
[15] K. T Kotani, Yoshimi, and M. Uchida, "Automatic Classification of Texts Written by Learners of English as a Foreign Language based on Linguistic Features and Learner Features," 2013, pp. 6305-6314.
[16] J. W. Pennebaker and L. A. King, ""Linguistic styles: Language use as an individual difference"," Journal of Personality and Social Psychology, 1999, vol. 77, pp. 1296-1312.
[17] J. W. Pennebaker, M. E. Francis, and R. J. Booth, "Linguistic Inquiry" Mahwah, NJ, USA: Erlbaum Publishers, 2001. (Online). http://www.erlbaum.com
[18] M. Coltheart, "The MRC psycholinguistic database," The Quarterly Journal of Experimental Psychology, 1981, vol. 33(4), pp. 497-505.
[19] M. Pennacchiotti and A.M. Popescu, "A Machine Learning Approach to Twitter User Classification," in ICWSM 11, 2011, pp. 281-288.
[20] B. L. Monroe, M. P. Colaresi, and K. M. Quinn, "Fighting words: Lexical feature selection and evaluation for identifying the content of political conflict," in Political Analysis, 2008, vol. 16(4), pp. 372-403.
[21] F. Heylighen and J. M. Dewaele, "Variation in the contextuality of language: an empirical measure", Context in Context, Special issue of Foundations of Science, 2002, vol. 7(3), pp. 293-340.
[22] M. R Mehl, S. D. Gosling, and J. W. Pennebaker, ""Personality in its natural habitat: Manifestations and implicit folk theories of personality in daily life", vol. 90, pp. 862-877, 2006.
[23] T. Yarkoni, "Personality in 100,000 Words: A Large-Scale Analysis of Personality and Word Use among Bloggers," National Institute of Health Public Access, 2010, pp. 1-23.
[24] C. Moral, A. d. Antonio, R. Imbert, and J. Ramirez, "A survey of stemming algorithms in information retrieval," in Information Research, 2014vol. 19(1).
[25] D. Maynard and A. Funk, "Automatic detection of political opinions in tweets," 2010, pp. 1-12.
[26] B. Liu, M. Hu, and J. Cheng, "Analyzing and comparing opinions on the web," in Proceedings of the 14th international conference on World Wide Web, 2005, pp. 342–351.