ESKİŞEHİR TECHNICAL UNIVERSITY JOURNAL OF SCIENCE AND TECHNOLOGY
A- APPLIED SCIENCES AND ENGINEERING

2018, 19(3), pp. 645 - 659, DOI: 10.18038/aubtda.412532

COMPARATIVE ANALYSIS OF RECENT FEATURE SELECTION METHODS FOR SENTIMENT CLASSIFICATION

Alper Kürşat UYSAL [1, *]

[1] Computer Engineering Department, Faculty of Engineering, Eskişehir Technical University, 26555, Eskişehir, TURKEY

ABSTRACT

Sentiment classification has become popular especially in recent years. As it is valid for all text classification problems, feature space’s high dimensionality is one of the biggest problems for sentiment classification due to accuracy considerations. This study analyses the performance of six recent text feature selection methods for document level sentiment classification using two widely-known classifiers namely Support Vector Machines (SVM) and naïve Bayes (NB). Three datasets including different types of sentiment data were utilized in the experiments. These datasets are named as Cornell movie review, Sentiment140, and Nine public sentiment. For evaluation, two different success measures namely Micro-F1 and Macro-F1 were used. Also, 3-fold cross-validation is preferred for a fair performance evaluation. Experiments indicated that distinguishing feature selector (DFS) and discriminative features selection (DFSS) methods are superior to the other four feature selection methods for sentiment classification. The highest classification performances with SVM classifier were obtained when it is combined with DFSS feature selection method in general. On the other hand, highest classification performances with NB classifier were obtained when it is combined with DFS feature selection method.

Keywords: Pattern recognition, Sentiment classification, Feature selection

1. INTRODUCTION

Sentiment classification has become popular as a result of the increase in online web platforms that people communicate. In recent years, people mostly prefer to share their opinions about many things on the Internet. These opinions may influence the behaviors of other people and effect their decisions on many things. For example, especially in recent years, most people check the reviews on the Internet before buying a product or watching a movie. These reviews include opinions of people about some products or movies. The aim of sentiment classification is to automatically classify opinions of people into some pre-defined categories. Opinions of people are generally stored as unstructured text documents and these pre-defined categories are either positive or negative in most cases. However, sentiment classification can be achieved on text-based review documents in three kind of levels [1] where they are called as document-level, sentence-level, and aspect-level. The aim of document-level sentiment classification is to extract sentiment degrees for text documents. These text documents can be anything such as movie review or Twitter message. Most of the publicly available datasets on the Internet are created for document-level sentiment classification. As another level for sentiment classification, sentence-level sentiment classification purposes detection of sentiments for separate sentences. Researchers benefit from parse trees in order to split the sentences into their syntactic units in some sentence-level sentiment classification studies [2]. Aspect level sentiment classification purposes to extract sentiment degrees for various entities addressed in reviews. For example, a product review may contain two kind of aspects such as price and quality but sentiment degrees related to these aspects may be different. There are many studies in the literature that deal with these three different types of sentiment classification levels. This study is focused on document-level sentiment classification. Thus, the studies related to this research field are taken into consideration. Recent studies and solutions about sentiment classification are briefly explained in the following paragraphs.

*Corresponding Author: akuysal@anadolu.edu.tr
Received: 04.04.2018 Accepted: 04.07.2018

As sentiment classification is one of the text classification domains, common text classification solutions can be applied besides such new approaches as deep learning [3, 4]. A common solution for text classification is to apply widely-known machine learning techniques after extracting features and constructing feature vectors. Bag-of-words [5] is the common approach which is used for feature extraction. The orders of terms are ignored in this approach and weighted frequencies of unique terms in the training set are used for representing text documents. Even a small collection may contain thousands of features and it is necessary to apply dimension reduction techniques after feature extraction in almost all cases. Feature selection, one of the widely-known dimension reduction techniques, is generally applied for text classification. Thus, feature selection becomes an essential step for text classification as it helps to reduce dimensionality and improve classification performance.

Feature selection is an on-going research field and so many recent methods are proposed especially for text classification. The studies which will be mentioned in the related work evaluate either earlier feature selection methods or feature selection methods which are not specifically proposed for text classification. Although feature selection methods proposed for pattern recognition problems can be applied on text classification domain, they are not optimized for solving text classification problems such as effectively dealing with high dimensionality. This study aims to analyze the effectiveness of recently proposed feature selection methods for classification of sentiments. In this study, the performances of six recent text feature selection methods in the literature are extensively analyzed for document level sentiment classification. It should be also noted that all of the feature selection methods considered as a part of this study are recent methods which are specifically proposed for text classification unlike the generalized feature selection methods proposed for pattern recognition such as ReliefF method. Support Vector Machines (SVM) and naïve Bayes (NB) classifiers are employed to test the efficacy of these approaches on three datasets including different types of sentiment data. Also, in the experiments, two different metrics namely Micro-F1 and Macro-F1 are used to measure success ratios. Deviation from Poisson distribution (DP), distinguishing feature selector (DFS), discriminative features selection (DFSS), improved comprehensive measurement feature selection (ICMFS), relative discrimination criterion (RDC), and feature selection based on interclass and intraclass relative contributions of terms (IIRCT) are feature selection methods utilized in this study. As a finding of experimental study, it can be said that DFS and DFSS feature selection methods offer better performance than the others. Rest of the paper is organized as follows: a literature review about related work is given in Section 2. Section 3 explains feature selection methods utilized in this study. Section 4 explains the classifiers employed in the experimental work in details. Section 5 describes the experimental settings and presents results obtained for each dataset. Finally, a conclusion is given in Section 6.

2. RELATED WORK

Document-level sentiment classification studies using the common text classification approaches are reviewed in this section. Saraee and Bagheri evaluated the effectiveness of various feature selection approaches for classification of Persian product reviews [6]. Term frequency variance (TFV), document frequency (DF), modified mutual information (MMI), and mutual information (MI) feature selection methods were used for this comparison. Then, the resulting feature vectors were fed into naïve Bayes (NB) classifier. They concluded that the success ratio obtained with MMI feature selection method is slightly better than the ones obtained with TFV and DF approaches. Mittal and Agarwal analyzed the impact of various types of features to classification performance of sentiments [7]. Unigram and bi-gram features obtained from text documents were used. Then, they analyzed the effectiveness of these features on four standard sentiment datasets. Minimum redundancy maximum relevancy (mRMR) and information gain (IG) methods were utilized for selecting features. They stated that mRMR is more successful than IG feature selection method for sentiment classification. Moreover, the success ratio obtained with multinomial naïve Bayes (MNB) classifier is better than the success ratio obtained with SVM classifier. Wang et al. evaluated four statistical feature selection methods namely document frequency (DF), chi square (CHI2), mutual information (MI), and information gain (IG) for sentiment classification of Chinese hotel reviews [8]. Adverbs, adjectives, and verbs were selected as possible features and SVM classifier was used for the evaluation. According to their findings, DF is the best performer and CHI2 is the runner-up. Omar et al. compared the performance of seven feature selection approaches namely information gain (IG), uncertainty, principal component analysis (PCA), Gini index, relief-f, chi square (CHI2), and Support Vector Machines for Arabic sentiment classification [9]. They used SVM, k-nearest neighbor (knn), and NB classifiers for evaluation. They stated the setting combining SVM-based feature selection and SVM classifier yields the best performance on Arabic sentiments. Akba et al. assessed feature selection methods for Turkish sentiment classification using NB and SVM classifiers [10]. They utilized information gain (IG) and chi square (CHI2) for selecting features in the experiments. According to their findings, SVM classifier slightly outperforms NB classifier and the performances of two feature selection methods are nearly identical. Prusa et al. analyzed the impact of ten different filter-based feature selection approaches for tweet sentiment classification using 4 classifiers namely knn, C4.5 decision tree, logistic regression (LR), and Multilayer Perceptron (MLP) [11]. The feature selection methods are chi-square (CHI2), Kolmogorov-Smirnov (KS) statistic, Gini index (GI), mutual information (MI), area under the precision-recall curve (PRC), probability ratio (PR), ROC curve, signal-to-noise (S2N) ratio, significance analysis of microarrays (SAM), and Wilcoxon rank sum (WRS). They concluded that feature selection can significantly increase classification performance for all classifiers. Uysal and Murphey compared the success ratios of the combination of three feature selection methods and SVM classifier with some deep learning based methods for sentiment classification [4]. They used Gini index (GI), information gain (IG), and distinguishing feature selector (DFS) for this comparison. They concluded that deep learning based methods generally surpassed feature selection based classification approaches. Besides, DFS and IG methods perform better than GI for selecting sentiment features. Onan and Korukoglu proposed an ensemble classification approach combining the feature sets obtained by some filter-based feature selection methods such as Gain ratio, information gain, symmetrical uncertainty coefficient, Chi-square, ReliefF, probabilistic significance measure, and Pearson correlation coefficient [12]. They utilized genetic algorithm in order to find a better performing feature set and they concluded that the proposed feature selection scheme is effective for sentiment classification.

3. FEATURE SELECTION METHODS

Filter-based feature selection approaches are mostly used for text classification in comparison to the other two alternatives namely wrapper and embedded methods. Filter-based methods execute faster than the others due to not including classifier interaction during feature selection process. Filter-based text feature selection methods, proposed between 2009 and 2016, are employed in this study. These methods are feature selection based on deviation from Poisson distribution (DP), distinguishing feature selector (DFS), discriminative features selection (DFSS), improved comprehensive measurement feature selection (ICMFS), relative discrimination criterion (RDC), and feature selection based on interclass and intraclass relative contributions of terms (IIRCT). Explanations about these methods are given in the following subsections.

3.1. Deviation from Poisson distribution (DP)

DP is a feature selection method derived from Poisson distribution [13]. The dependency between the feature and the corresponding class is low if the feature fits in Poisson distribution. In this case, the feature score is small and the feature is considered to be less discriminative. On the other hand, the feature is considered to be more discriminative when the feature score is greater. DP score of a feature for each class can be calculated using Eq. 1.

$$DP(t,C) = \frac{(a-\hat{a})^2}{\hat{a}} + \frac{(b-\hat{b})^2}{\hat{b}} + \frac{(c-\hat{c})^2}{\hat{c}} + \frac{(d-\hat{d})^2}{\hat{d}}$$

$$\hat{a} = n(C)\{1-\exp(-\lambda)\}$$

$$\hat{b} = n(C)\exp(-\lambda)$$

$$\hat{c} = n(\bar{C})\{1-\exp(-\lambda)\}$$

$$\hat{d} = n(\bar{C})\exp(-\lambda) \qquad (1)$$

$$\lambda = \frac{F}{N},$$

In Eq. 1, N is the total amount of text documents in the training data and F is the total of frequencies of feature t in entire collection. $n(C)$ and $n(\bar{C})$ indicate the sum of text documents labelled as class C and labelled as other classes except class C, respectively. λ symbol expresses the expected frequency of the feature t. The quantities a and b represent the amount of text documents including feature t and the amount of text documents not including feature t in all text documents of class C, respectively. c is the amount of text documents including feature t but not belonging to class C. However, d is the amount of text documents not belonging to class C and not including feature t simultaneously. Finally, $\hat{a},\hat{b},\hat{c},\hat{d}$ are predicted versions of a, b, c, d, respectively. After calculating class-based feature scores, these scores are globalized using weighted averaged globalization function [14] to obtain a unique DP score for each feature.

3.2. Distinguishing feature selector (DFS)

DFS is a popular and well-performing filter-based feature selection technique for text classification [14]. DFS relies on some pre-defined criteria. DFS score of a feature can be calculated using Eq. 2.

$$DFS(t) = \sum_{i=1}^{M} \frac{P(C_i \mid t)}{P(\bar{t} \mid C_i) + P(t \mid \bar{C}_i) + 1} \qquad (2)$$

In this equation, $P(C_i \mid t)$ denotes to the conditional probability of class C_i given presence of term t and M denotes to the total count of classes. However, $P(\bar{t} \mid C_i)$ denotes to the conditional probability of lack of term t given class C_i and $P(t \mid \bar{C}_i)$ denotes to the conditional probability of term t given all other classes except C_i.

3.3. Discriminative features selection (DFSS)

DFSS is one of the recently proposed filter-based feature selection methods for text classification [15]. It aims to select features with a higher average term frequency and a higher document frequency in documents of a certain class. DFSS score of a feature for each class can be calculated using Eq. 3.

$$DFSS(t,C) = \frac{\mathrm{tf}(t,C) / \mathrm{df}(t,C)}{\mathrm{tf}(t,\overline{C}) / \mathrm{df}(t,\overline{C})} \times \frac{a}{(a+b)} \times \frac{a_i}{(a+c)} \times \left| \frac{a}{(a+b)} - \frac{c}{(c+d)} \right| \tag{3}$$

In this equation, $\mathrm{tf}(t,C)$ and $\mathrm{tf}(t,\overline{C})$ are the frequencies of feature t occurring in category C and occurring in other categories, respectively. However, $\mathrm{df}(t,C)$ denotes to the amount of text documents feature t occurs in category C and $\mathrm{df}(t,\overline{C})$ denotes to the amount of text documents feature t occurs in other categories. a is the amount of text documents in category C containing feature t and b denotes to the amount of text documents in category C not including feature t. c is the amount of text documents in all categories except C that contains feature t. d is the amount of text documents in all categories except C which are not containing feature t. After calculating class-based feature scores, these scores are globalized using maximum globalization function.

3.4. Improved Comprehensive Measurement Feature Selection (ICMFS)

ICMFS is an extended version of existing comprehensive measurement feature selection method [16]. ICMFS score of a feature for each class can be calculated using Eq. 4.

$$\mathrm{ICMFS}(t,\mathrm{C}) = \frac{P(t \mid \mathrm{C}) \times P(C \mid t)}{P(\mathrm{C}) \times \max_{\mathrm{C}_j \neq \mathrm{C}} \left\{ \dfrac{P(t \mid \mathrm{C}_j) + \Delta}{P_{avg} + \Delta} \right\}} \tag{4}$$

In the formula, Δ is a constant that the authors proposing ICMFS method set to 0.001. P_{avg} denotes to the average of $P(t \mid \mathrm{C}_j)$ scores when $\mathrm{C}_j \neq \mathrm{C}$. After calculating class-based feature scores, these scores are globalized using maximum globalization function.

3.5. Relative discrimination criterion (RDC)

RDC is a new feature ranking metric, which considers document frequencies for each term count of a term [17]. RDC takes into account the difference between document frequencies for corresponding term counts of a term in the positive and negative classes. It is not a probabilistic feature selection method like most of the other filter-based approaches. The flow of RDC algorithm can be given as below.

POS = amount of text documents belonging to positive class
NEG = amount of text documents belonging to negative class
TCMAX = maximum term count for term t
for tc = 1 to TCMAX do
 tp_{tc} *= amount of positive text documents including term t having term count tc*
 fp_{tc} *= amount of negative text documents including term t having term count tc*
 $tpr_{tc} = tp_{tc} / POS$
 $fpr_{tc} = fp_{tc} / NEG$
 $D_{tc} = | tpr_{tc} - fpr_{tc} |$
 $$RDC_{tc} = \frac{D_{tc}}{\min(\mathrm{tpr_{tc}}, \mathrm{fpr_{tc}}) * \mathrm{tc}}$$
end
$AUC_{tc} = 0$
for tc = 1 to TCMAX do

$$AUC_{tc} = AUC_{tc} + \frac{RDC_{tc} + RDC_{tc+1}}{2}$$

end

After calculating class-based feature scores, these scores are globalized using weighted average globalization function.

3.6. Feature selection based on interclass and intraclass relative contributions of terms (IIRCT)

IIRCT [18] is a feature selection method considering interclass and intraclass contributions of terms. IIRCT is formulated as below.

$$IIRCT(t, \mathrm{C}_k) = \sum_{j=1,\, j\neq k}^{M} \left[P(t \mid \mathrm{C}_k) \times P(C_k \mid t) - P(t \mid \mathrm{C}_j) \times P(C_j \mid t) \right] \tag{5}$$

After calculating class-based feature scores, these scores are globalized using maximum globalization function.

4. CLASSIFIERS

Two widely-known classifiers are utilized to examine the contributions of the selected features to the performance of classification. These classifiers are linear support vector machine (SVM) classifier [19] and naïve Bayes (NB) classifier [20]. These classification algorithms are proven to be successful for text classification [5, 21, 22]. The statements in the next subsections explain these two classifiers.

4.1. Support vector machines (SVM)

SVM classifier aims to maximize the margin between two classes [19, 23]. SVM aims finding a decision hyperplane whose distance is maximum from data points located along the axis belonging to two classes. It is required to identify data points that are located closer to the border between two classes and these points are named as support vectors. Therefore, the resulting classifier is called as SVM classifier. SVM can be a linear or nonlinear classifier according to kernel parameter. Linear SVM is employed in this study as it is very successful for text classification.

4.2. Naïve Bayes (NB)

NB is an example to probabilistic classifiers which are based on Bayes theorem. They assume that there exist an independence between features. Therefore, in naïve Bayes classification, a probability score is determined with multiplication of the conditional probabilities. One of the widely-used event models for naïve Bayes is Gaussian event model. However, multi-variate Bernoulli and multinomial event models are known as successful event models which are specific to text classification [24]. Multi-variate Bernoulli and Multinomial event models differs in calculation of a probability in their formula. Multi-variate Bernoulli event model is used in this study for implementation of naïve Bayes classifier.

5. EXPERIMENTAL WORK

A comprehensive analysis was realized in this section to compare the performances of six filter-based feature selection methods. Term weighting is realized using term frequency inverse document frequency method (TF-IDF). The characteristics including distributions of utilized datasets are explained, the success measures are introduced and the results of experiments are presented in the following subsections.

5.1. Datasets

Three datasets with various characteristics are used in this study. The first one is Cornell movie review dataset [25] which is a widely known sentiment analysis dataset employed in many previous studies. The second one is a partition of Sentiment140 dataset [26]. The third dataset is Nine public sentiment dataset including positive and negative review documents from camera, camp, doctor, drug, laptop, TV, lawyer, music, and radio domains [27]. For all of three datasets, three fold cross-validation is used for a fair evaluation. More information about these datasets is given in Tables 1-3.

Table 1. Cornell movie review dataset

Class Label	Number of Documents
positive	1000
negative	1000

Table 2. Sentiment140 dataset

Class Label	Number of Documents
positive	26921
negative	23079

Table 3. Nine public sentiment dataset

Class Label	Number of Documents
positive	3018
negative	3016

5.2. Success Measures

Micro-F1 and Macro-F1 [28, 29] success measures are utilized in this study. Micro-F1 score is calculated disregarding class distribution of documents. Therefore, all classification results in the entire dataset are taken into consideration. When the classes in dataset are skewed, Micro-F1 scores can be favored by large classes. Micro-F1 measure can be expressed as

$$Micro-F1 = \frac{2 \times p \times r}{p + r}, \tag{6}$$

In this formula, p and r correspond to precision and recall for all decisions of the corresponding classifier. Macro-F1 score is calculated for each class in the corresponding dataset. Then, the scores of all classes are averaged. Therefore, classes are equally weighted unlike Micro-F1. Macro-F1 measure can be expressed as

$$Macro-F1 = \frac{\sum_{k=1}^{C} F_k}{C}, \qquad F_k = \frac{2 \times p_k \times r_k}{p_k + r_k}, \tag{7}$$

In this formula, p_k and r_k correspond to precision and recall scores belonging to class k, respectively.

5.3. Accuracy Analysis

In this section, the ratio of correct classification decisions of six filter-based text feature selection methods are analyzed using Micro-F1 and Macro-F1 scores. The selected features for each setting are sent to SVM and NB classifiers as input. In the experiments, different feature sizes are used. Lowercase conversion and stemming [30] are utilized as the two pre-processing steps in addition to weighting terms with TF-IDF. Micro-F1 and Macro-F1 scores obtained in the experiments are listed in Figures 1-12 for each dataset, respectively.

According to Figures 1-4, highest Micro-F1 and Macro-F1 scores for SVM classifier were obtained using DFSS feature selection method on Cornell Movie dataset. On the other hand, highest Micro-F1 and Macro-F1 scores for NB classifier were obtained using DFS feature selection method. However, the perfomance of IIRCT feature selection method is worse than the others for Cornell Movie dataset. It should be also noted that the performance of NB classifier is better than SVM classifier for Cornell Movie dataset according to the highest Micro-F1 and Macro-F1 scores.

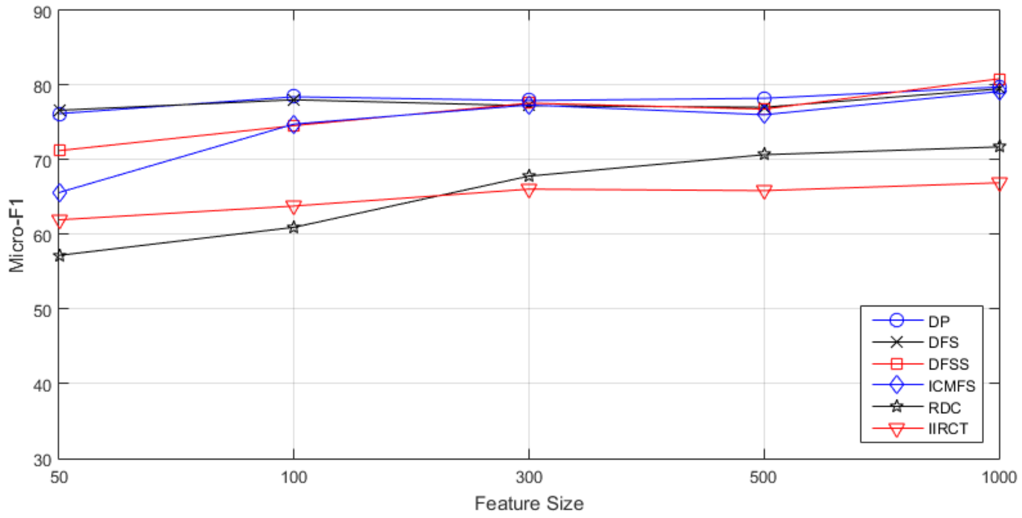

Figure 1. Micro-F1 scores (%) for Cornell Movie dataset using SVM

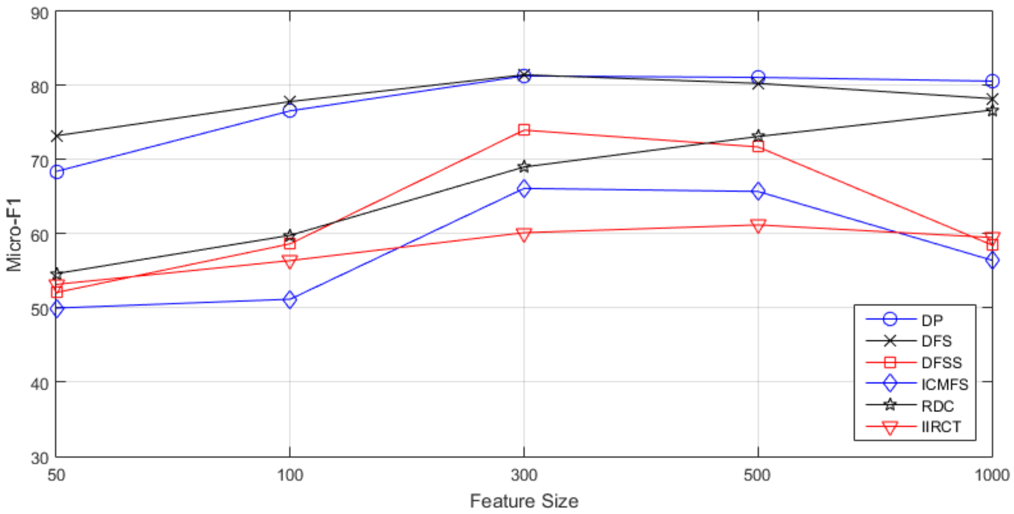

Figure 2. Micro-F1 scores (%) for Cornell Movie dataset using NB

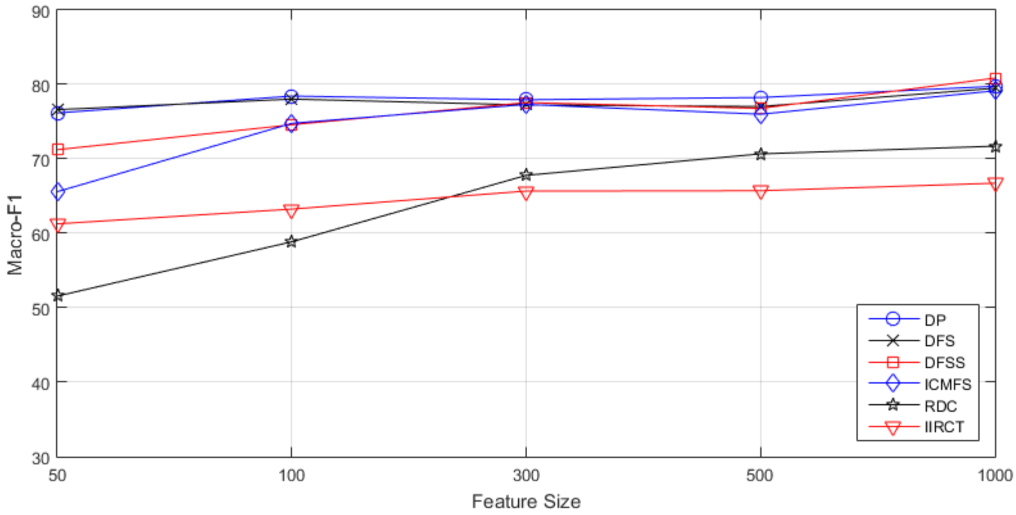

Figure 3. Macro-F1 scores (%) for Cornell Movie dataset using SVM

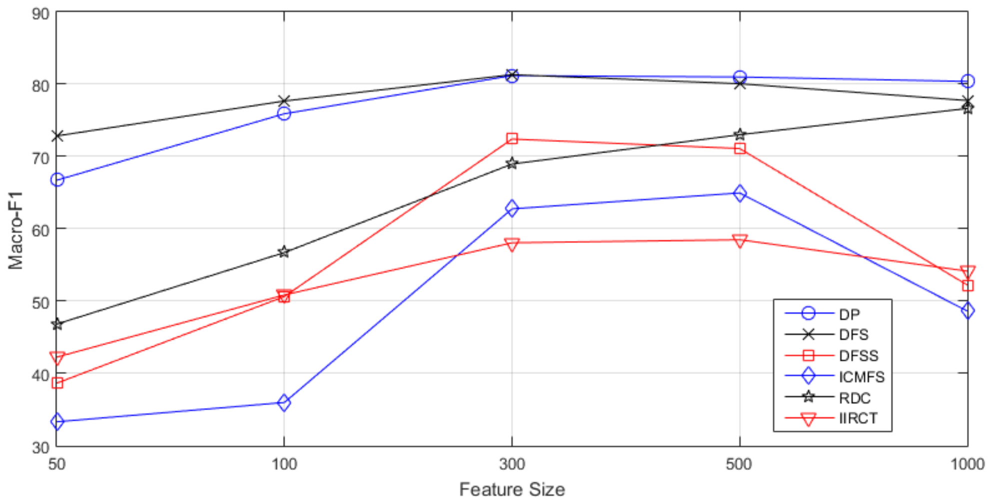

Figure 4. Macro-F1 scores (%) for Cornell Movie dataset using NB

According to Figures 5-8, highest Micro-F1 and Macro-F1 scores for SVM classifier were attained using DFSS feature selection method on Sentiment140 dataset. On the other hand, highest Micro-F1 and Macro-F1 scores for NB classifier were obtained using DFS feature selection method. The perfomance of IIRCT feature selection method is worse than the others for Sentiment140 dataset. It should be also noted that the performance of SVM is better than NB classifier for Sentiment140 dataset according to the highest Micro-F1 and Macro-F1 scores. As Sentiment140 dataset contains tweets which are very short text documents, the highest performances are obtained with highest feature sizes.

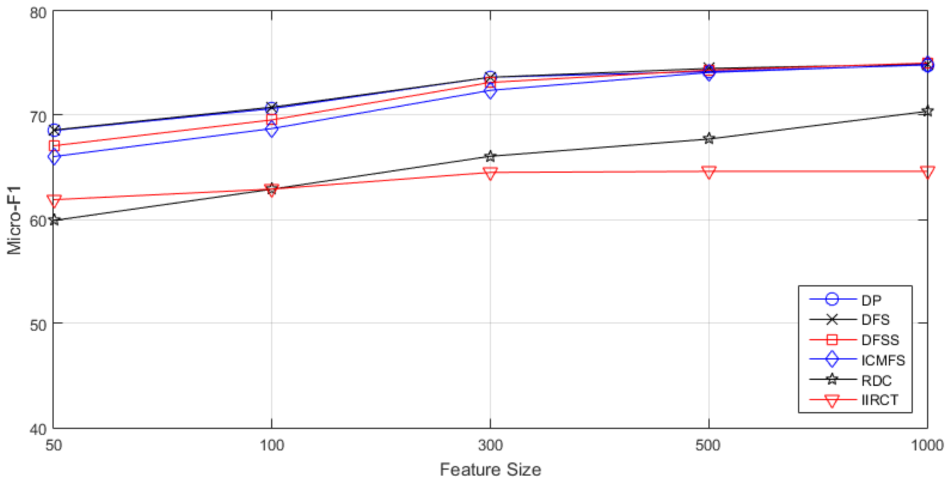

Figure 5. Micro-F1 scores (%) for Sentiment140 dataset using SVM

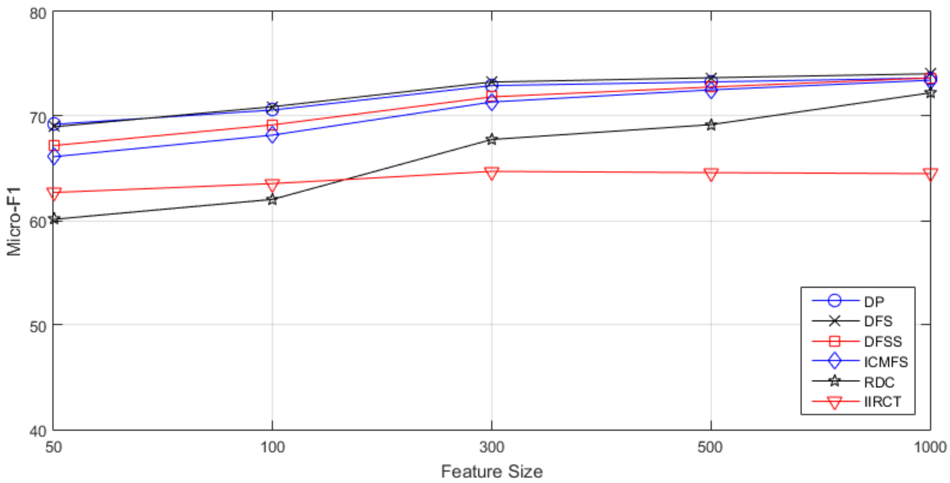

Figure 6. Micro-F1 scores (%) for Sentiment140 dataset using NB

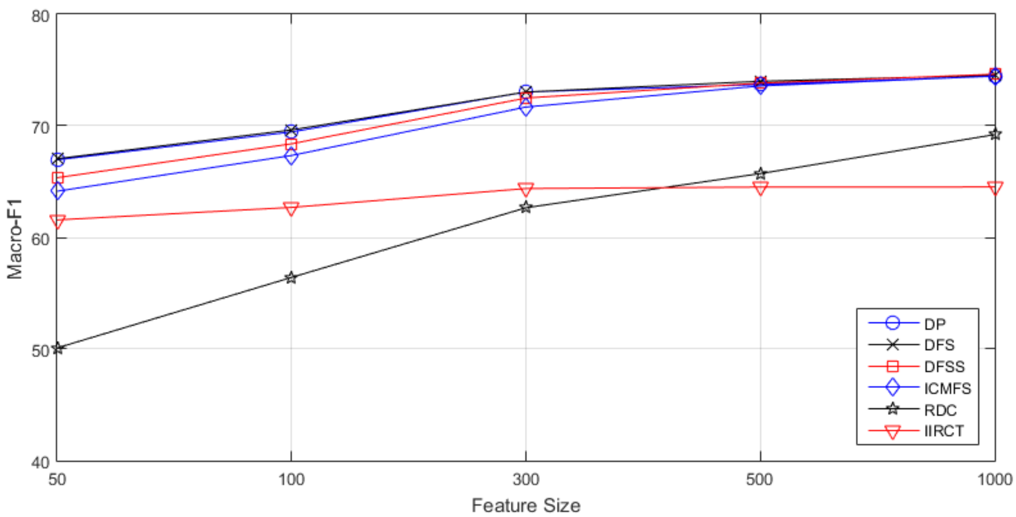

Figure 7. Macro-F1 scores (%) for Sentiment140 dataset using SVM

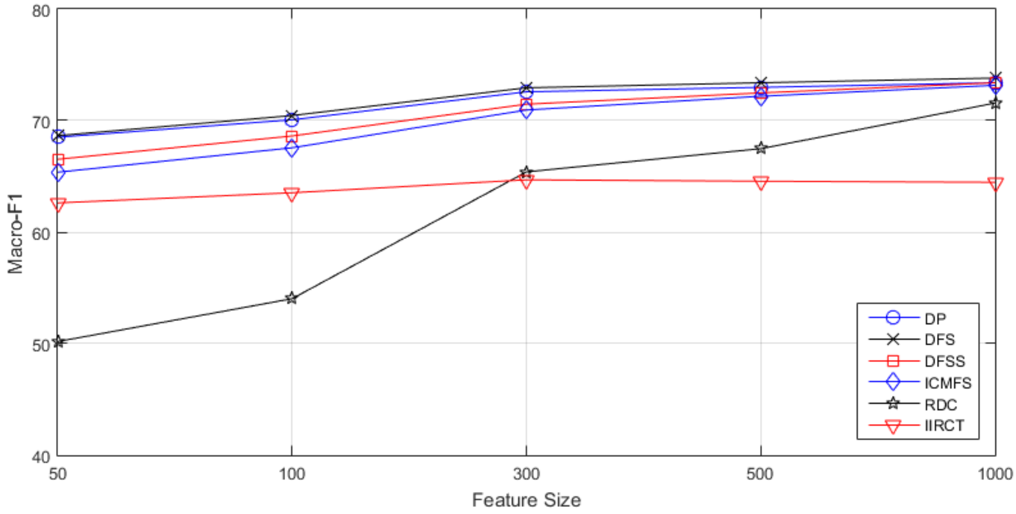

Figure 8. Macro-F1 scores (%) for Sentiment140 dataset using NB

According to Figures 9-12, highest Micro-F1 and Macro-F1 scores for SVM classifier were obtained using DP feature selection method on Nine public sentiment dataset. On the other hand, highest Micro-F1 and Macro-F1 scores for NB classifier were achieved using DFS feature selection method. However, the perfomance of IIRCT feature selection method is worse than the others for Sentiment140 dataset as it is valid for the other sentiment datasets. It should be also noted that the performance of SVM is better than NB classifier for Nine public sentiment dataset according to the highest Micro-F1 and Macro-F1 scores.

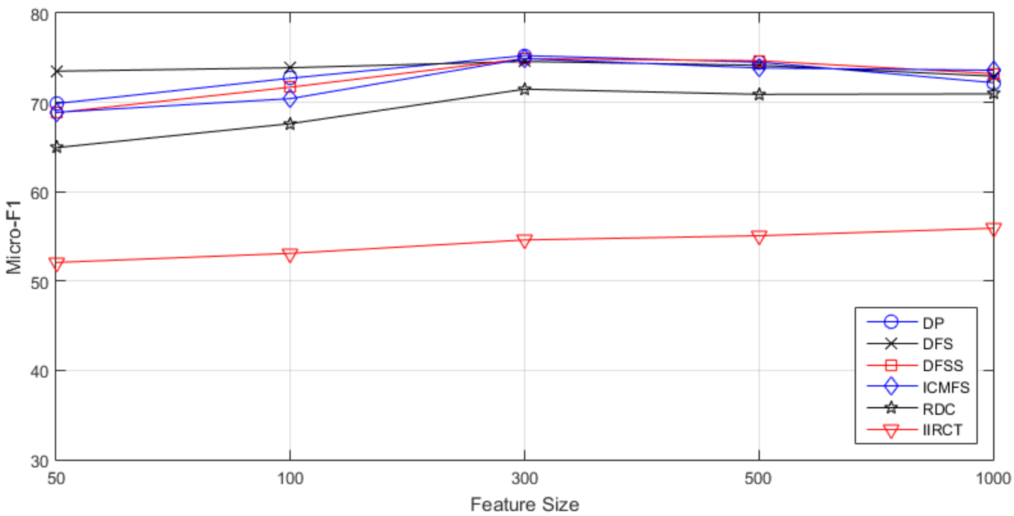

Figure 9. Micro-F1 scores (%) for Nine public sentiment dataset using SVM

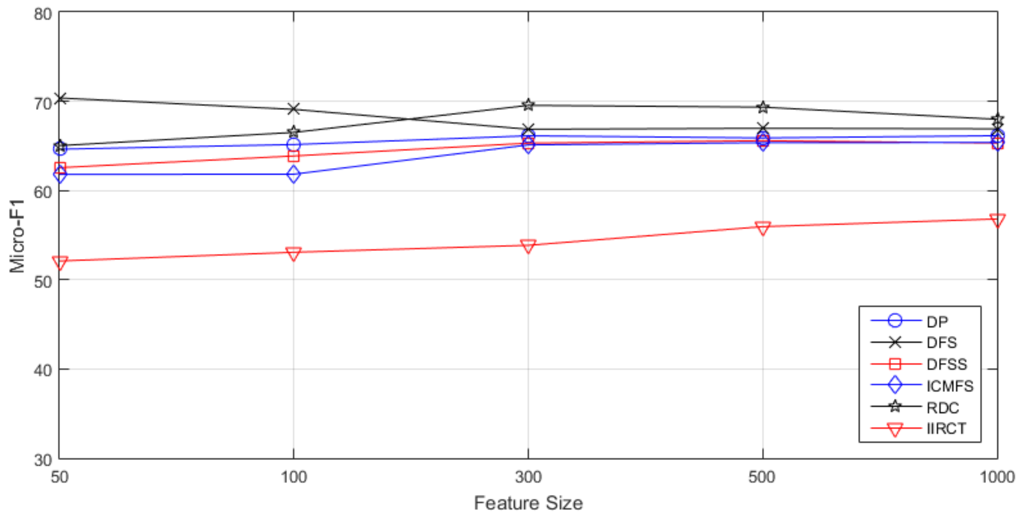

Figure 10. Micro-F1 scores (%) for Nine public sentiment dataset using NB

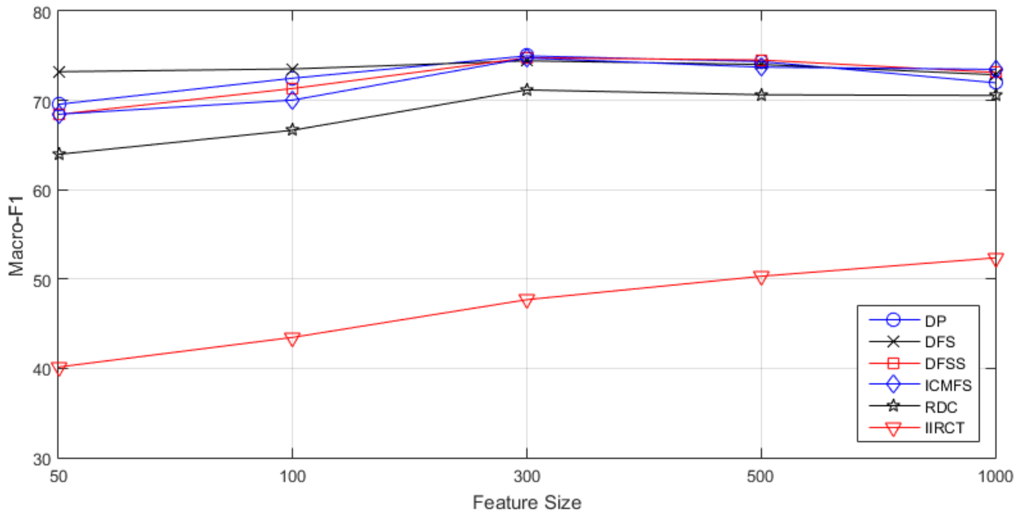

Figure 11. Macro-F1 scores (%) for Nine public sentiment dataset using SVM

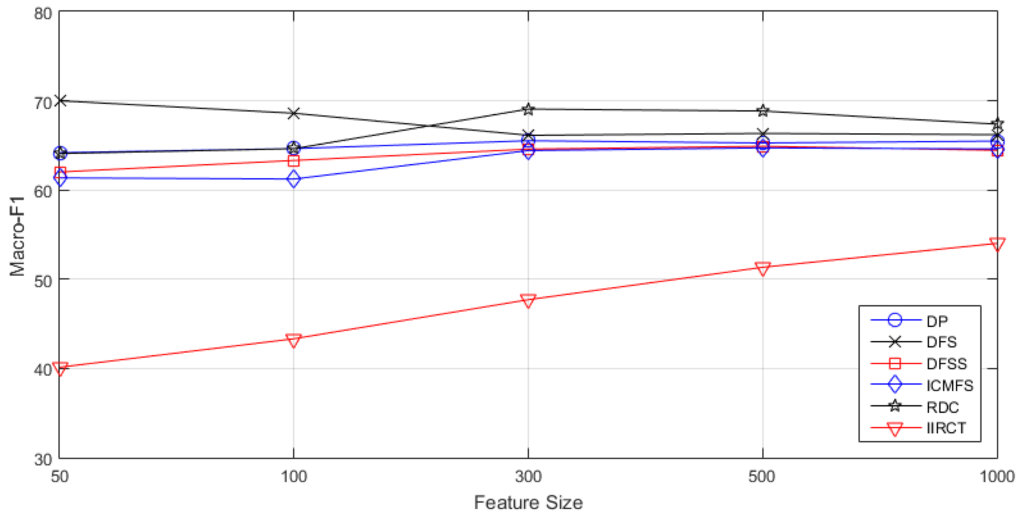

Figure 12. Macro-F1 scores (%) for Nine public sentiment dataset using NB

When overall highest Micro-F1 and Macro-F1 scores are considered, DFS performed better than the others for 6 out of 12 cases. The runner-up is DFSS which obtains better performances than the others for 4 out of 12 cases. However, DP performed better for 2 out of 12 cases. It should be noted that ICMFS, RDC, and IIRCT didn't obtain the better scores for any cases.

SVM classifier obtained the better scores on two datasets and NB outperformed SVM classifier on only Cornell Movie dataset. Highest scores with SVM classifier were obtained when it is combined with DFSS feature selection method in general. On the other hand, highest scores with NB classifier were obtained when it is combined with DFS feature selection method for all cases.

6. CONCLUSIONS

In this study, the performance of six recent text feature selection methods in the literature are examined for document level sentiment classification. SVM and NB classifiers are utilized to test the efficacy of these approaches. Three datasets including different types of sentiment data are utilized in the experiments. These datasets are named as Cornell movie review, Sentiment140, and Nine public sentiment. Cornell movie dataset and Nine public sentiment dataset include relatively long reviews in comparison to Sentiment140 dataset because Sentiment140 dataset is constructed using tweets. In this study, two different success measures namely Micro-F1 and Macro-F1 are used. Experiments indicated that DFS and DFSS feature selection methods are superior to the other feature selection methods for sentiment classification. The highest accuracies were obtained with the combination of DFS feature selection and naïve Bayes classifier for Cornell Movie dataset. However, the highest accuracies were obtained with the combination of DFSS feature selection and SVM classifier for Sentiment140 and Nine public sentiment datasets. As a future work, the performances of these recent text feature selection methods can be analyzed for different text classification domains.

REFERENCES

[1] Feldman R. Techniques and applications for sentiment analysis. Communications of the ACM 2013; 56: 82-89.

[2] Socher R, Perelygin A, Wu JY, Chuang J, Manning CD, Ng AY, et al. Recursive deep models for semantic compositionality over a sentiment treebank, in Proceedings of the conference on empirical methods in natural language processing (EMNLP); 2013, pp. 1631-1642.

[3] Araque O, Corcuera-Platas I, Sánchez-Rada JF, Iglesias CA. Enhancing deep learning sentiment analysis with ensemble techniques in social applications. Expert Systems with Applications 2017; 77: 236-246.

[4] Uysal AK, Murphey YL. Sentiment classification: Feature selection based approaches versus deep learning, in 17th IEEE International Conference on Computer and Information Technology (CIT); 2017, pp. 23-30.

[5] Uysal AK. An improved global feature selection scheme for text classification. Expert Systems with Applications 2016; 43: 82-92.

[6] Saraee M, Bagheri A. Feature selection methods in Persian sentiment analysis, in 18th International Conference on Applications of Natural Language to Information Systems; Salford, UK; 2013, pp. 303-308.

[7] Agarwal B, Mittal N. Optimal feature selection for sentiment analysis, in Computational Linguistics and Intelligent Text Processing: 14th International Conference (CICLing 2013); Samos, Greece; 2013, pp. 13-24.

[8] Wang H, Yin P, Yao J, Liu JNK. Text feature selection for sentiment classification of Chinese online reviews. Journal of Experimental & Theoretical Artificial Intelligence 2013; 25: 425-439.

[9] Omar N, Albared M, Al-Moslmi T, Al-Shabi A. A comparative study of feature selection and machine learning algorithms for Arabic sentiment classification, in 10th Asia Information Retrieval Societies Conference (AIRS 2014); Kuching, Malaysia; 2014, pp. 429-443.

[10] Akba F, Uçan A, Sezer E, Sever H. Assessment of feature selection metrics for sentiment analyses: Turkish movie reviews, in 8th European Conference on Data Mining; 2014, pp. 180-184.

[11] Prusa JD, Khoshgoftaar TM, Dittman DJ. Impact of feature selection techniques for tweet sentiment classification, in FLAIRS Conference; 2015, pp. 299-304.

[12] Onan A, Korukoğlu S. A feature selection model based on genetic rank aggregation for text sentiment classification. Journal of Information Science 2017; 43: 25-38.

[13] Ogura H, Amano H, Kondo M. Feature selection with a measure of deviations from Poisson in text categorization. Decision Support Systems 2009; 36: 6826-6832.

[14] Uysal AK, Gunal S. A novel probabilistic feature selection method for text classification. Knowledge-Based Systems 2012; 36: 226-235.

[15] Zong W, Wu F, Chu L-K, Sculli D. A discriminative and semantic feature selection method for text categorization. International Journal of Production Economics 2015; 165: 215-222.

[16] Feng L, Zuo W, Wang Y. Improved comprehensive measurement feature selection method for text categorization, in Network and Information Systems for Computers (ICNISC), 2015 International Conference on; 2015, pp. 125-128.

[17] Rehman A, Javed K, Babri HA, Saeed M. Relative discrimination criterion – A novel feature ranking method for text data. Expert Systems with Applications 2015; 42: 3670-3681.

[18] Zhou H, Guo J, Wang Y, Zhao M. A feature selection approach based on interclass and intraclass relative contributions of terms. Computational Intelligence and Neuroscience 2016; 1-8.

[19] Joachims T. Text categorization with support vector machines: Learning with many relevant features, in 10th European Conference on Machine Learning; Chemnitz, Germany; 1998, pp. 137-142.

[20] Chen J, Huang H, Tian S, Qu Y. Feature selection for text classification with Naïve Bayes. Expert Systems with Applications 2009; 36: 5432-5435.

[21] Kumar MA, Gopal M. A comparison study on multiple binary-class SVM methods for unilabel text categorization. Pattern Recognition Letters 2010; 31: 1437-1444.

[22] Yu B, Zhu DH. Combining neural networks and semantic feature space for email classification. Knowledge-Based Systems 2009; 22: 376-381.

[23] Theodoridis S, Koutroumbas K. Pattern Recognition, 4 ed., Academic Press, 2008.

[24] Jiang L, Cai Z, Zhang H, Wang D. Naive Bayes text classifiers: A locally weighted learning approach. Journal of Experimental & Theoretical Artificial Intelligence 2013; 25: 273-286.

[25] Pang B, Lee L. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts, in Proceedings of the 42nd annual meeting on Association for Computational Linguistics; 2004, pp. 271.

[26] Go A, Bhayani R, Huang L. Twitter sentiment classification using distant supervision. CS224N Project Report, Stanford 2009; 1: 1-6.

[27] Whitehead M, Yaeger L. Building a general purpose cross-domain sentiment mining model, in World Congress on Computer Science and Information Engineering; 2009, pp. 472-476.

[28] Manning CD, Raghavan P, Schutze H. Introduction to Information Retrieval. New York, USA, Cambridge University Press, 2008.

[29] Gunal S. Hybrid feature selection for text classification. Turkish Journal of Electrical Engineering and Computer Sciences 2012; 20: 1296-1311.

[30] Porter MF. An algorithm for suffix stripping. Program 1980; 14: 130-137.