

Article

Ensemble Based Classification of Sentiments Using Forest Optimization Algorithm

Mehreen Naz ¹, Kashif Zafar ¹ and Ayesha Khan ²

¹ Department of Computer Science; National University of Computer and Emerging Sciences, Lahore 54770, Pakistan; mehru.naz@gmail.com

² School of Science and Technology; University of Management and Technology, Lahore 54782, Pakistan; ayesha.khan@umt.edu.pk

* Correspondence: kashif.zafar@nu.edu.pk

Received: 8 April 2019; Accepted: 21 May 2019; Published: 23 May 2019



Abstract: Feature subset selection is a process to choose a set of relevant features from a high dimensionality dataset to improve the performance of classifiers. The meaningful words extracted from data forms a set of features for sentiment analysis. Many evolutionary algorithms, like the Genetic Algorithm (GA) and Particle Swarm Optimization (PSO), have been applied to feature subset selection problem and computational performance can still be improved. This research presents a solution to feature subset selection problem for classification of sentiments using ensemble-based classifiers. It consists of a hybrid technique of minimum redundancy and maximum relevance (mRMR) and Forest Optimization Algorithm (FOA)-based feature selection. Ensemble-based classification is implemented to optimize the results of individual classifiers. The Forest Optimization Algorithm as a feature selection technique has been applied to various classification datasets from the UCI machine learning repository. The classifiers used for ensemble methods for UCI repository datasets are the k-Nearest Neighbor (k-NN) and Naïve Bayes (NB). For the classification of sentiments, 15–20% improvement has been recorded. The dataset used for classification of sentiments is Blitzer’s dataset consisting of reviews of electronic products. The results are further improved by ensemble of k-NN, NB, and Support Vector Machine (SVM) with an accuracy of 95% for the classification of sentiment tasks.

Keywords: feature subset selection; classification; ensemble; evolutionary algorithms; data mining; sentiment analysis

1. Introduction

Classification of sentiments is basically a technique to determine the polarity of a given text, document, or sentence. Sentiment analysis and classification use both machine learning and natural language processing (NLP) techniques. In today’s era of the Internet, all of us are using social media as a regular source of information on a daily basis. Social websites have become an important source of information. People share their opinions about almost everything, e.g., any product, book, movie, social or political issues, etc., on these websites. Mostly these reviews are in the form of text. Publicly shared reviews over blogs or articles are used to identify the unanimous customer opinion for any product or service to help maintain good consumerism [1]. The major issue that arises while gathering the data from social networking sites is that the reviews mostly contain grammatical and/or spelling mistakes and data is usually so large that correcting those mistakes is humanly impossible. Whenever the data is being extracted from any of the social networking sites it usually contains large parts of unwanted information, including html tags, as compared to actual meaningful and useful information comprising of review text. There are several pre-processing techniques to apply on the extracted data first and

then it is analyzed. Feature Subset Selection (FSS) is used as a pre-processing technique in data mining for high dimensional data. In the case of social media mining, it is normal for the data to be analyzed to have even higher dimensions. Classification of such data with reasonable computational cost has become an important topic of research in recent years. Most of the solutions proposed for sentiment analysis are either based on classifiers or dataset pre-processing techniques to improve classification accuracy. Genetic Algorithm (GA) and Particle Swarm Optimization (PSO) have been used for feature subset selection from the Artificial Intelligence domain. The GA- and PSO-based solutions have improved classification accuracy but these solutions are computationally expensive, so it affects the overall performance. GA and PSO are meta-heuristic algorithms that use population of initial solutions. They are used for optimization problems. This research presents an ensemble based classification of sentiments using a recently developed evolutionary algorithm, named the Forest Optimization Algorithm (FOA) and proposed in [2]. The seeding procedure of trees is simulated in this algorithm. It is inspired by some trees that outlive other trees based on their better survival conditions. FOA produces a best tree (subset of features) among all other trees based on performance. FOA has outperformed GA and PSO when applied to benchmark functions and an optimization problem of feature weighting using continuous weights [2].

2. Literature Review

Feature selection helps reduce number of attributes to be stored in database to get rid of irrelevant or redundant features to produce more useful and efficient results. Feature subset selection algorithms are either categorized as filter and wrapper approaches or as complete search, heuristic search, meta-heuristic search methods, and methods based on artificial neural networks (ANN) [3]. There are two approaches for feature selection; the filter approach was one of the earliest techniques used for feature subset selection. The filter approach makes use of characteristics and statistics of data instead of learning algorithms. Mostly, it performs two methods of ranking the variables by evaluating the importance of them and subset selection of feature [4]. The second approach is the Wrapper approach that makes use of learning algorithms while searching for a suitable feature subset. The chosen classifier is a part of feature selection procedure; wrapper method makes good feature selection by considering the classification accuracy as a part of its evaluation function [4]. Wrapper methods, when compared to filter methods, are more time consuming for datasets with large numbers of features. On the other hand, they do produce better results and are more accurate than filter methods. There are researchers who have tried to solve sentiments analysis by using different techniques from machine learning domain and by using ensemble-based classification techniques [5–12]. In [7], authors present the use of sentiment analysis as a technique for analyzing the presence of human trafficking in escort ads pulled from the open web. Traditional techniques have not focused on sentiment as a textual cue of human trafficking and instead have focused on other visual cues (e.g., the presence of tattoos in associated images), or textual cues (specific styles of ad-writing; keywords, etc.). They applied two widely cited sentiment analysis models: the Netflix and Stanford models, and also train binary and categorical (multiclass) sentiment models using escort review data crawled from the open web. The individual model performances and exploratory analysis motivated them to construct two ensemble sentiment models that correctly serve as a feature proxy to identify human trafficking 53% of the time when evaluated against a set of 38,563 ads provided by the DARPA MEMEX project. In [11], authors explore the effects of feature selection on sentiment analysis of Chinese online reviews. Firstly, N-char-grams and N-POS-grams are selected as the potential sentimental features. Then, the improved Document Frequency method is used to select feature subsets, and the Boolean Weighting method is adopted to calculate feature weight. The chi-square test is carried out to test the significance of experimental results. The results suggest that sentiment analysis of Chinese online reviews obtains higher accuracy when taking 4-POS-grams as features. Furthermore, the improved document frequency achieves significant improvement in sentiment analysis of Chinese online reviews.

The mRMR (minimum Redundancy and Maximum Relevance)-based [13] feature selection method first focuses on the correlation among the features. The features must not be highly correlated to each other as it may cause redundancy. Secondly, the relevance of features with the class label is taken into account. Mutual information (MI) is the way to measure the level of similarity or simply the redundancy between features. To measure the maximum relevance of features with respect to the class again we use mutual information. We need MI for the relevance measure to be maximized and MI for the redundancy measure to be minimized [14]. Two methods are designed to combine the redundancy and relevance of features in one function: (1) MID: Mutual Information Difference criterion and (2) MIQ: Mutual Information Quotient criterion. The divisive combination of redundancy and relevance, i.e., MIQ has outperformed the other difference method in the case of discrete data [14]. They have applied these mRMR criteria for gene selection task.

Sentiment classification is basically a task of opinion mining used to extract people's unanimous reviews about any topic, event, or product [1]. Sentiment analysis or opinion mining is a step-wise procedure to produce the results of this classification task. Usually this classification is binary, either the reviews are positive or negative about the respective topic, event, or product. Mostly, sentiment analysis (SA) and opinion mining (OM) are interchangeable but Walaa Medhat in 2014 defined them with slight differentiation [15]. Sentiment analysis is used when the sentiment expressed in some document or text is to be analyzed, whereas opinion mining is used to extract people's opinion for analysis. Sentiment analysis analyzes the text's sentiment and then identifies its sentiment polarity. The three-step procedure for sentiment analysis given in [16], includes (i) corpus collection, (ii) corpus analysis, and (iii) training the classifier. It has mainly focused on the collection of data for analysis because there is still a lack of benchmark datasets for sentiment classification problem. Most of the datasets are based on reviews taken from micro-blogging websites, like IMDB and Amazon.com, etc. IMDB has movie reviews while Amazon has reviews for a wide variety of products. The most important and earliest step in sentiment classification is feature extraction or selection where some features from the text are selected to analyze the sentiment of the chosen text or document.

Two of the most famous and frequently used evolutionary algorithms are GA and PSO. The main advantage of using evolutionary algorithms is that they are robust and easily adaptable to changing circumstances. They are even more efficient when combined with other optimization techniques. Evolutionary algorithms are widely applicable and are known to provide solutions for problems which other techniques have either failed to solve or have given less efficient solutions [17]. Genetic algorithm has been applied to feature selection problem with different variations and hybridization with other algorithms/methods, successfully over the years. Ahmed Abbasi, Hsinchun Chen, and Arab Salem used GA to select a features-based information gain technique and then used those features for sentiment classification of English and Arabic text [18]. A hybrid method was proposed which uses an ensemble of GA and Naïve Bayes for sentiment classification [17]. It has improved the accuracy of movie review dataset to 93.80%. Kalaivani used hybrid of information gain based genetic algorithm with bagging technique for feature selection for opinion mining. This GA based hybrid method has shown high accuracy of 87.50% when applied to movie review dataset [16]. The Particle Swarm Optimization (PSO) algorithm has been applied to the problem of feature selection for sentiment analysis in recent years. Most classifiers have combined this algorithm with other existing machine learning classifying techniques and have succeeded to achieve better classification accuracy. Bernali Sahu and Debahuti Mishra [19] presented a novel feature selection algorithm based on PSO for cancer microarray data. This novel algorithm for feature selection outperformed K-NN, SVM, and PNN (Probabilistic Neural Network) [19]. There are approaches based on deep learning for the sentiment analysis problem. The results are comparable with other techniques but computational cost is comparatively high as compared to conventional approaches [20].

A hybrid of PSO-SVM was presented by Basari et al. in 2013; this hybrid method helped improve the sentiment classification accuracy from 71.87% to 77% for the IMDB movie review dataset [21]. In 2014 another hybrid approach of PSO and ACO (Ant Colony Optimization) was proposed for

web-based opinion mining by George Stylios. This bio-inspired method achieved 90.59% accuracy using 10 fold cross validation approach while outperforming the C4.5 algorithm with 83.66% accuracy [1]. In 2014 PSO was also used to select features for thermal face recognition and results showed a success rate of 90.28% for all images [22]. Lin Shang et al. has used binary version of PSO after modifying the basic algorithm into sentiment classification oriented approach and applied it to customer review datasets. The results have shown that this is superior to the basic binary PSO-based feature selection algorithm [23]. Ensemble-based classifiers are called multi-classifier systems (MCS). MCS give better results than implementing individual classifiers alone [24]. The basic topologies to design MCS are explored below.

2.1. Conditional Topology

Conditional topology usually works with one classifier selected as a primary classifier out of all the available classifiers in a multi-classifier system or an ensemble. If the first classifier fails to correctly classify the input data it then goes to the second classifier. The selection of the next classifier can either be static or dynamic. One example of dynamic selection can be decision trees. The process of classification can continue until the data is correctly classified or if all the classifiers have been used. MCS can be computationally efficient if the primary classifier is efficient. One way to keep it computationally efficient is to keep the heavy classifiers at the end of the queue of available classifiers [25].

2.2. Hierarchical (Serial) Topology

This topology of classifiers help narrow down the most accurate classification as the data passes through the classifiers. The classifiers are used in succession where, with each classifier, the error of classification gets minimized or the result gets more focused on the actual class of data. The classifier with minimum error should be the successor here. Although, the correct classification must be ensured after every classifier, otherwise, the next classifier will not be able to correctly classify the input data [25].

2.3. Hybrid Topology

As some classifiers perform better on certain types of datasets, thus, the hybrid topology basically selects the best classifier for the type of input data [25]. To select the best classifiers different type of classifiers are implemented in this multi-classifier mechanism or an ensemble.

2.4. Multiple (Parallel) Topology

Parallel or multiple topology based system of multi-classifiers is the most commonly used. In this system, all the classifiers operate in parallel on the input data. The result of each classifier is integrated in one place or function, then the final classification is chosen according to the implemented design logic to select the appropriate classifier for the data type [25].

3. Proposed Methodology

The feature subset selection process is performed by Forest Optimization Algorithm and pre-processing of features or attributes has been done by mRMR technique. The mRMR technique will eliminate irrelevant and redundant features before applying the feature subset selection module. The flow chart of the proposed methodology is given in Figure 1. The proposed approach combines mRMR- and FOA-based feature subset selection to achieve efficient and better solutions. The flow of this proposed approach is shown here:

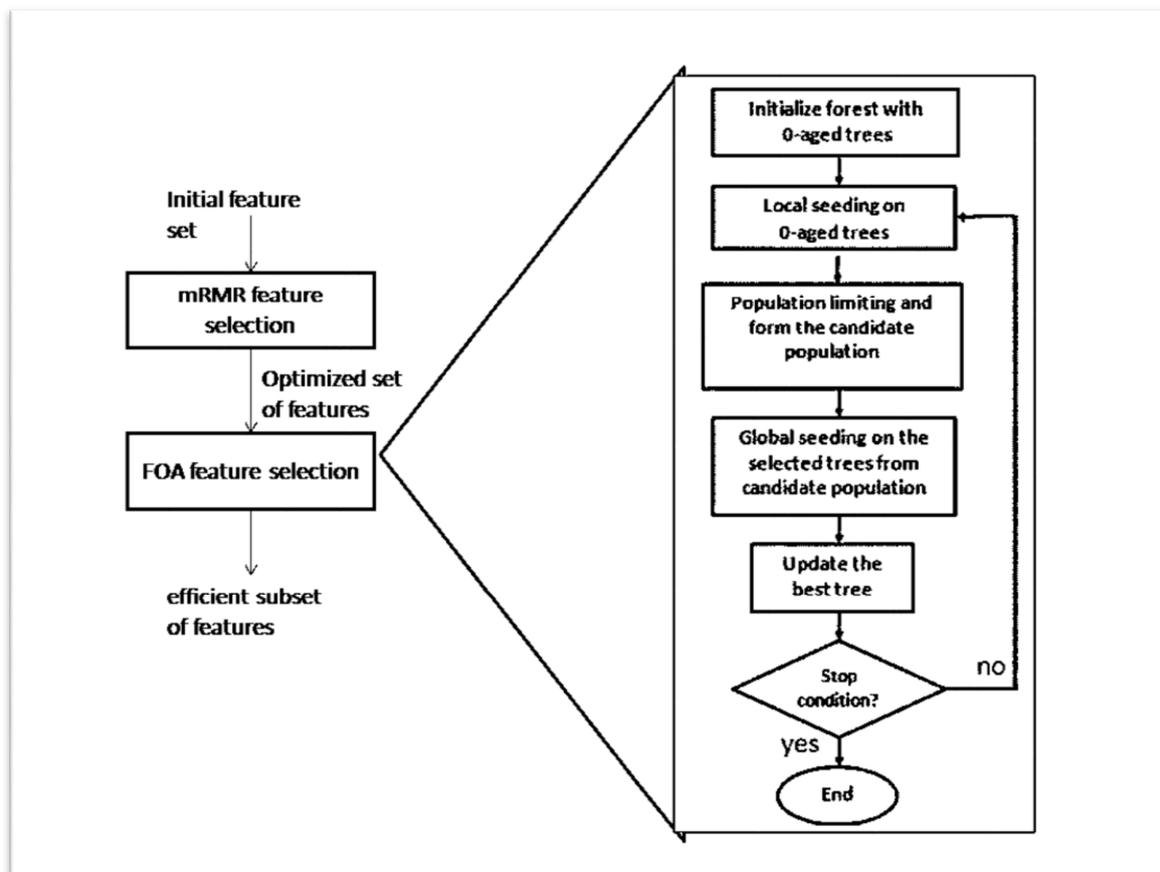


Figure 1. Flowchart of mRMR-based FOA.

3.1. Step 1: Preprocessing of sentiment analysis datasets

Dataset

The dataset for sentiment analysis we have used for this classification task is taken from Amazon and constructed by John Blitzer and Mark Dredze [26]. This dataset consists of different kinds of products and their reviews, including electronics, kitchen appliances, books, and DVDs. All reviews are in raw form with html tags of review text, review id, date, title, rating, product, and user location. Product reviews with star ratings have five stars. Star rating reviews are converted to positive reviews if they have more than three stars or negative if they have less than three stars, and the rest of the reviews have been discarded because of their ambiguous polarity. We have used electronics and book review datasets for the classification task. To analyze these reviews, preprocessing was required. Following are the preprocessing steps that have been applied:

- Review id, review text, and label were extracted from html tags. Review text is the basic text that we need for analysis.
- All text was converted to lower case and punctuation signs were removed.
- Stop words were removed.
- Alphanumeric characters were removed.
- Stemming was applied by using Porter Stemmer method [27].
- Words of length less than three letters were also removed.

The input to algorithm must be in a form of feature vector. After preprocessing, we get the bag of words (BOW). This BOW basically forms our feature vector. The preprocessing step has been implemented in MATLAB.

3.2. Step 2: mRMR- and Forest Optimization Algorithm-Based Feature Subset Selection

The bag-of-words or feature vector that we have created in step 1 is very large in size and it still contains many unnecessary features. The unnecessary and irrelevant features affect the classification accuracy and the larger size of feature vector increases the computational time. To avoid this we need to perform feature subset selection methods on this feature vector to achieve a more useful and efficient set of features.

3.2.1. mRMR Feature Subset Selection

mRMR (minimum Redundancy and Maximum Relevance) is used to reduce the size of feature vector [5]. The subset of features obtained through mRMR is efficient and produces more accurate results. Results and experiments have shown that the MIQ (Mutual Information Quotient) criterion is more suitable to use with discrete data out of the two mRMR criteria. The set of features obtained by using MIQ, are then forwarded to Forest Optimization Algorithm to obtain the best subset of features. This subset of features will give the highest classification accuracy.

Minimal Redundancy and Maximal Relevance (mRMR) is based on the mutual information that measures the amount of information shared between two features. The mutual information between two variables A and B is defined as the joint probability distribution of the two variables A and B.

The mutual information of a feature and a class is referred to as relevance of the feature with the class. The mutual information of a feature with other features is referred to as redundancy of the feature. mRMR for any feature k is expressed as follows:

$$mRMR_k = \frac{Relevance_k}{Redundancy_k}$$

This measure of mRMR emphasizes that low redundancy and high relevance depicts that the feature is mutually exclusive to other features and highly dependent on the class [5].

3.2.2. Forest Optimization Algorithm

Forest Optimization Algorithm (FOA) has been used for feature subset selection. It is inspired by the seeding process of trees in a forest [2]. There are trees in forest that live up to decades and some only survive for a limited period of time. A tree's survival depends on the conditions in the area in which they have been planted. FOA simulates natural seed dispersion process, which are basically of two kinds, local and global seed dispersion processes. In the local seed distribution, the seeds just fall under the trees and begin to grow, whereas in global seed distribution seeds were carried away to places that are far-away through animals, winds, or flowing water. Local seed distribution is referred as "local seeding" and global seed distribution as "global seeding" in FOA.

The initial population of trees is a matrix in which each row represents a potential solution and known as a "tree". One tree is $1 \times N_{var}$ dimensional. Keeping track of each tree's age makes it $1 \times (N_{var} + 1)$ dimensional vector. This $(N + 1)$ dimensional vector is our feature vector.

The basic steps involved in the Forest Optimization Algorithm are explored below.

3.2.3. Trees Initialization

Initial population of trees is randomly generated with age 0. The age limit of trees is the predefined parameter "Lifetime" for every tree. When any tree outgrows the age limit it will be omitted from the candidate tree population. The first variable of each tree represents its age and other variables are the features of a solution vector. Initially all features are present in the vector. Feature presence is represented by 1 and absence by 0 in Table 1.

Table 1. Tree encoding.

<i>Age</i>	v_1	v_2	...	$v_{N_{var}}$
0	1	1	...	1

Local Seeding

The function of local seeding is only applied to the newly generated trees with age 0 as shown in Figures 2 and 3. The value of local seeding operator is also a predefined parameter here called “Local seeding changes (LSC)”. This LSC operator tells how many new trees will be generated in result of applying local seeding function on each tree. Random numbers equal to LSC are generated within the range of 1 to N_{var} . If $LSC = 2$, we have $N = 4$, and random numbers are 2 and 3, then the values of feature number 2 and 3 are changed to form the new neighboring tree. This is repeated LSC number of times on each tree. Initially local seeding will be applied to all the trees in the forest. In further iterations number of newly generated trees will decrease because the age of all trees except the new ones will be incremented.

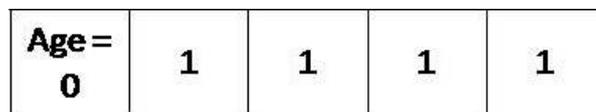


Figure 2. Tree before local seeding.

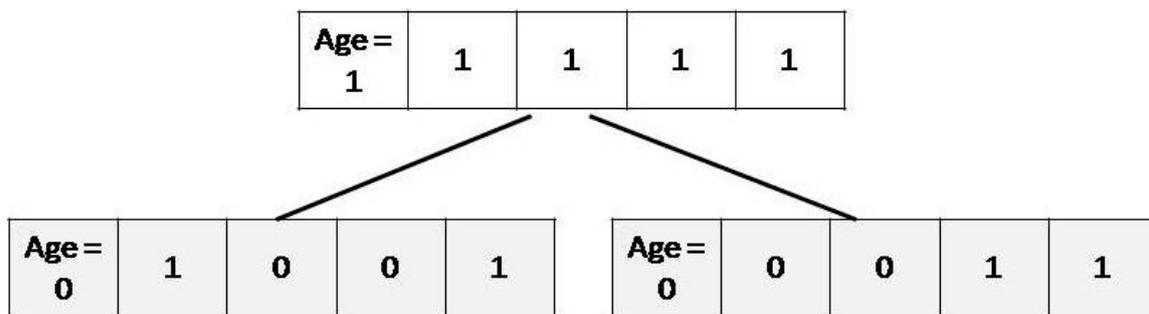


Figure 3. Tree after local seeding.

Population Limiting

Population limiting is needed to avoid infinite expansion of the forest. The parameter used to define the number of trees is called “area limit”. Area limit is defined equal to the initial number of trees in the forest [2]. Population limiting involves two steps; first, remove all the trees that have crossed their age limit, i.e., life-time, and add them to the candidate population. Secondly, sort the trees according to their fitness values in descending order, which in our case is the classification accuracy. After sorting if the number of trees is greater than the area limit, remove the extra trees and add them to the candidate population.

Global Seeding

Global seeding is applied on some percentage of the candidate population. This percentage is the “transfer rate” parameter. The “Global Seeding Changes (GSC)” is another predefined parameter. Global seeding is applied on selected trees and in result the value of GSC number of features of each selected tree is changed. These changes result into a new tree with age 0. This newly generated tree is added to the forest.

Updating the Best Tree

Trees are sorted according to their fitness value or classification accuracy. The age of best tree is updated to '0' so that the age limit parameter would not rule it out. As local seeding is only applied to the trees with age '0', this step helps in locally optimizing the best solution that we have so far.

Stopping Conditions

Three stop conditions can be considered like all evolutionary algorithms:

- Fixed number of iterations
- No change observed in fitness value for number of iterations
- Required level of fitness value is achieved

3.3. Step 3: Ensemble-Based Classification

Three basic classifiers are used for ensemble model of sentiment classification. k -Nearest Neighbor (k -NN), Naïve Bayes (NB), and Support Vector Machine (SVM). All these supervised machine learning classifying algorithms are used in parallel topology of multi-classifier system (ensemble). In our proposed approach, parallel classifiers will optimize the classification accuracy for the sentiment analysis task.

4. Results and Analysis

Initially, the Forest Optimization Algorithm is first applied to some benchmark classification datasets and their results are compared with individual classifiers like Naïve Bayes and k -Nearest Neighbor. All datasets are downloaded from the UCI Machine Learning Repository. The details of the used datasets are shown below in Table 2.

Table 2. Benchmark classification datasets from UCI repository.

Names	No. of Instances	No. of Attributes	No. of Classes	Type of Attributes
Iris	150	4	3	Real
Bank note authentication	1372	4	2	Real
Teaching Assistant Evaluation	151	5	3	Categorical, Integer
Hepatitis	155	19	2	Categorical, Integer, Real
Image Segmentation	210	19	7	Real
Parkinson's Disease	195	23	2	Real
Ionosphere	351	34	2	Integer, Real
Dermatology	366	34	6	Categorical, Integer
Lung Cancer	32	56	3	Integer

4.1. Classification without FOA

Classification accuracy achieved from Naïve Bayes and k -nearest neighbors on the above mentioned datasets is shown here in Table 3. k -NN is used with the value of $k = 1$.

Table 3. Classification accuracy of NB and k -NN ($k = 1$) on UCI datasets.

Names	No. of Instances	No. of Attributes	No. of Classes	k -NN ($k = 1$)	NB
Iris	150	4	3	95.30%	96.00%
Bank note authentication	1372	4	2	54.50%	49.10%
Teaching Assistant Evaluation	151	5	3	62.90%	55.60%
Hepatitis	155	19	2	78.00%	81.20%
Image Segmentation	210	19	7	87.10%	77.60%
Parkinson's Disease	195	23	2	98.40%	39.40%
Ionosphere	351	34	2	86.30%	82.60%
Dermatology	366	34	6	94.00%	97.00%
Lung Cancer	32	56	3	34.30%	53.10%

4.2. Classification with FOA

Classification accuracy achieved from FOA is shown in Table 4. Features are selected using simple FOA here. FOA gives promising results on these benchmark datasets. The results with FOA and without FOA are compared in Table 5. The number of features reduced by FOA is better as compared to original features along with improving the classification accuracy.

Table 4. Classification accuracy using FOA on UCI datasets.

Names	No. of Instances	No. of Attributes	No. of Classes	<i>k</i> -NN (<i>k</i> = 1)	NB
Iris	150	4	3	96.00%	96.00%
Bank note authentication	1372	4	2	99.00%	87.30%
Teaching Assistant Evaluation	151	5	3	63.50%	52.00%
Hepatitis	155	19	2	88.30%	86.40%
Image Segmentation	210	19	7	88.30%	87.60%
Parkinson's Disease	195	23	2	100.00%	92.30%
Ionosphere	351	34	2	92.00%	92.30%
Dermatology	366	34	6	98.90%	97.50%
Lung Cancer	32	56	3	40.60%	59.30%

Table 5. Comparison of classification accuracy with and without FOA on UCI datasets.

Names	<i>k</i> -NN (<i>k</i> = 1)w/o FOA	NB w/o FOA	<i>k</i> -NN (<i>k</i> = 1) with FOA	NB with FOA
Iris	95.30%	96.00%	96.00%	96.00%
Bank note authentication	54.50%	49.10%	99.00%	87.30%
Teaching Assistant Evaluation	62.90%	55.60%	63.50%	52.00%
Hepatitis	78.00%	81.20%	88.30%	86.40%
Image Segmentation	87.10%	77.60%	88.30%	87.60%
Parkinson's Disease	98.40%	39.40%	100.00%	92.30%
Ionosphere	86.30%	82.60%	92.00%	92.30%
Dermatology	94.00%	97.00%	98.90%	97.50%
Lung Cancer	34.30%	53.10%	40.60%	59.30%

4.3. mRMR Based Feature Subset Selection Using FOA

mRMR feature subset selection is applied to all these datasets to improve classification results with less no. of features. Datasets with their original number of features and selected features using mRMR technique are shown in Table 6.

Table 6. Number of features reduced using mRMR on UCI datasets.

Names	No. of Instances	No. of Attributes	No. of Classes	mRMR Attributes	NB	<i>k</i> -NN (<i>k</i> = 1)
Hepatitis	155	19	2	5	82.58%	77.41%
Image Segmentation	210	19	7	10	77.61%	87.61%
Parkinson's Disease	195	23	2	10	97.43%	100%
Ionosphere	351	34	2	20	89.17%	87.46%
Dermatology	366	34	6	20	98.08%	97.26%
Lung Cancer	32	56	3	30	75.00%	56.25%

The abovementioned number of features are selected based on their improved classification accuracy. Classification accuracy is achieved by using two different initial random populations of 30 and 50 and shown in Tables 7 and 8.

Table 7. Classification accuracy of mRMR-FOA with a population size of 30 on UCI datasets.

Names	No. of Instances	Attributes in Dataset	mRMR Selected Attributes	mRMR-FOA Selected Attributes	No. of Runs	k -NN ($k = 1$)	NB
Hepatitis	155	19	5	5	50 100	87.09%	84.51%
Image Segmentation	210	19	13	10	50 100	89.52%	62.85%
Parkinson's Disease	195	23	10	8	50 100	100.00%	99.49%
Ionosphere	351	34	20	16	50 100	94.01%	89.17%
Dermatology	366	34	20	16	50 100	98.63%	97.81%
Lung Cancer	32	56	30	25	50 100	50.00%	62.50%

Table 8. Classification accuracy of mRMR-FOA with a population size of 50 on UCI datasets.

Names	No. of Instances	Attributes in Dataset	mRMR Selected Attributes	mRMR-FOA selected Attributes	No. of Runs	k -NN ($k = 1$)	NB
Hepatitis	155	19	5	5	50 100	87.09%	84.51%
Image Segmentation	210	19	13	10	50 100	89.52%	62.85%
Parkinson's Disease	195	23	10	8	50 100	100.00%	99.49%
Ionosphere	351	34	20	17	50 100	94.01%	89.17%
Dermatology	366	34	20	16	50 100	98.63%	97.81%
Lung Cancer	32	56	30	25	50 100	37.50% 43.70%	62.50% 68.75%

Comparative analysis has been performed on benchmark datasets to compare results of FOA + mRMR with competitor algorithms like PSO and GA. Table 9 shows result of Naïve Bayes classifier on three evolutionary algorithms used with mRMR technique [13]. The results obtained for lung cancer and image segmentation datasets are improved by 6% and 10%, respectively.

Table 9. Comparison of FOA with GA and PSO for NB on UCI datasets.

Names	No. of Features in Dataset	GA + mRMR		PSO + mRMR		FOA + mRMR	
		Accuracy	No. of Attributes	Accuracy	No. of Attributes	Accuracy	No. of Attributes
Hepatitis	19	83.90%	10	89.30%	10	84.51%	5
Image Segmentation	19	71.1%	9	73.1%	9	62.85%	10
Ionosphere	34	91.70%	16	92.30%	18	89.17%	16
Dermatology	34	94.90%	22	96.30%	20	97.81%	16
Lung Cancer	56	84.40%	25	88.10%	25	68.75%	25

4.4. mRMR Based Sentiment Classification Using FOA

For sentiment classification, a product review dataset taken from Amazon and constructed by John Blitzer and Mark Dredze [14] has been used. Reviews for four different types of products they have gathered include books, DVD, electronic, and kitchen appliances. It covers sentiment classification results of Blitzer's electronic products' reviews dataset. A total of 7028 dimensions of electronic products' reviews dataset are extracted after pre-processing. These 7028 attributes are reduced to 1200 after the application of mRMR (minimum Redundancy and Maximum Relevance) technique. The classification accuracy achieved without applying mRMR was less than 80%. The improved results can be seen in Tables 10 and 11.

For sentiment classification we have used three classifiers with FOA, k -NN, NB, and SVM. We have used the dataset with three different number of instances to observe the performance of FOA with the increase in number of instances.

Table 10. Sentiment classification using mRMR-FOA for the products' reviews dataset (population size 50).

No. of Instances	Attributes in Dataset	mRMR-FOA Selected Attributes	No. of Runs	k -NN ($k = 1$)	NB	SVM
100	7028	1200	10	77.00%	81.00%	93.00%
			20	82.00%	87.00%	95.00%
150	7028	1200	10	80.67%	76.67%	90.00%
			20	84.00%	82.00%	92.00%
200	7028	1200	10	80.00%	79.00%	-
			20	87.50%	86.00%	-

Table 11. Sentiment classification using mRMR-FOA for the products' reviews dataset (population size 30).

No. of Instances	Attributes in Dataset	mRMR-FOA Selected Attributes	No. of Runs	k -NN ($k = 1$)	NB	SVM
100	7028	1200	10	90.00%	81.00%	90.00%
			20	90.00%	87.00%	94.00%
150	7028	1200	10	79.67%	76.00%	87.33%
			20	83.00%	78.67%	92.00%

4.5. Results from Ensemble Of Classifiers

Ensemble of three classifiers i.e., k -NN, NB, and SVM, optimizes the output by picking out the result with highest classification accuracy. In all the experiments of sentiment classification task SVM has performed the best out of the three mentioned classifiers. The results can be seen in Table 12.

Table 12. Results from ensemble of classifiers for products' reviews dataset.

Populations	Attributes in Dataset	mRMR-FOA Selected Attributes	k -NN ($k = 1$)	NB	SVM
30	7028	1200	87.50%	87.00%	95.00%
50		1200	90.00%	88.00%	94.00%

5. Discussion and Conclusion

The hybrid of FOA- k -NN and FOA-NB has outperformed single k -NN and NB classifiers when applied to various benchmark classification datasets downloaded from the UCI repository. After these improved results, the proposed approach of mRMR-based feature subset selection with FOA was applied on the product review dataset taken from Amazon. For sentiment classification we have implemented three classifiers, including SVM, to observe the classification accuracy. The experimentation was carried out on three different numbers of instances with two sizes of initial population. The increase in the size of instances gradually increase the computation time of FOA. Out of all three classifiers, SVM takes more time with FOA, whereas the computation time of k -NN and NB gradually increase with the number of instances. The sentiment classification accuracy of this dataset has improved 15–20% when the FOA-mRMR feature subset selection technique is applied. The ensemble of k -NN, NB, and SVM classifiers is implemented to optimize the final output. The optimized final result for two different populations of 30 and 50 is 95% and 94%, respectively.

Author Contributions: Formal analysis, A.K.; methodology, M.N. and K.Z.; validation, A.K.; writing—original draft, M.N.; writing—review and editing, A.K.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Stylios, G.; Katsis, C.D.; Christodoulakis, D. Using Bio-inspired intelligence for Web opinion Mining. *Int. J. Comput. Appl.* **2014**, *87*, 36–43. [[CrossRef](#)]

2. Ghaemi, M.; Feizi-Derakhshi, M.R. Forest optimization algorithm. *Expert Syst. Appl.* **2014**, *41*, 6676–6687.
3. Feizi-Derakhshi, M.R.; Ghaemi, M. Classifying different feature selection algorithms based on the search strategies. In Proceedings of the International Conference on Machine Learning, Electrical and Mechanical Engineering, Dubai, United Arab Emirates, 8–9 January 2014.
4. Feature Selection. Available online: https://en.wikipedia.org/wiki/Feature_selection (accessed on 2 January 2017).
5. Halim, Z.; Atif, M.; Rashid, A.; Edwin, C.A. Profiling players using real-world datasets: Clustering the data and correlating the results with the big-five personality traits. *IEEE Trans. Affect. Comput.* **2017**. [[CrossRef](#)]
6. Mensikova, A.; Mattmann, C.A. Ensemble Sentiment Analysis to Identify Human Trafficking in Web Data. Available online: http://www.hrl.com/laboratories/issl/ccni/workshop/gta3/papers/GTA3_paper_5.pdf (accessed on 15 January 2017).
7. Halim, Z.; Waqas, M.; Baig, A.R.; Rashid, A. Efficient clustering of large uncertain graphs using neighborhood information. *Int. J. Approx. Reason.* **2017**, *90*, 274–291. [[CrossRef](#)]
8. Halim, Z.; Muhammad, T. Quantifying and optimizing visualization: An evolutionary computing-based approach. *Inf. Sci.* **2017**, *385*, 284–313. [[CrossRef](#)]
9. Zheng, L.; Wang, H.; Gao, S. Sentimental feature selection for sentiment analysis of Chinese online reviews. *Int. J. Mach. Learn. Cybern.* **2018**, *9*, 75–84. [[CrossRef](#)]
10. Muhammad, T.; Halim, Z. Employing artificial neural networks for constructing metadata-based model to automatically select an appropriate data visualization technique. *Appl. Soft Comput.* **2016**, *49*, 365–384. [[CrossRef](#)]
11. Hu, Z.; Hu, J.; Ding, W.; Zheng, X. Review sentiment analysis based on deep learning. In Proceedings of the 2015 IEEE 12th International Conference on e-Business Engineering, Beijing, China, 23–25 October 2015.
12. Cambria, E. Affective computing and sentiment analysis. *IEEE Intell. Syst.* **2016**, *31*, 102–107. [[CrossRef](#)]
13. Huang, C.; Zhu, J.; Liang, Y.; Yang, M.; Fung, G.P.C.; Luo, J. An efficient automatic multiple objectives optimization feature selection strategy for internet text classification. *Int. J. Mach. Learn. Cybern.* **2019**, *10*, 1151–1163. [[CrossRef](#)]
14. Peng, H.; Long, F.; Ding, C. Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 1226–1238. [[CrossRef](#)] [[PubMed](#)]
15. Kalaivani, P.; Shunmuganathan, K.L. Feature reduction based on genetic algorithm and hybrid model for opinion mining. *Sci. Program.* **2015**, *2015*, 12. [[CrossRef](#)]
16. Pak, A.; Paroubek, P. Twitter as a corpus for sentiment analysis and opinion mining. In Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC), Valletta, Malta, 19–21 May 2010; pp. 1320–1326.
17. Govindarajan, M. Sentiment analysis of movie reviews using hybrid method of naive bayes and genetic algorithm. *Int. J. Adv. Comput. Res.* **2013**, *3*, 139.
18. Abbasi, A.; Chen, H.; Salem, A. Sentiment analysis in multiple languages: Feature selection for opinion classification in web forums. *ACM Trans. Inf. Syst.* **2008**, *26*, 12. [[CrossRef](#)]
19. Sahu, B.; Mishra, D. A novel feature selection algorithm using particle swarm optimization for cancer microarray data. *Procedia Eng.* **2012**, *38*, 27–31. [[CrossRef](#)]
20. Chachra, A.; Mehndiratta, P.; Gupta, M. Sentiment analysis of text using deep convolution neural networks. In Proceedings of the 2017 Tenth International Conference on Contemporary Computing (IC3), NOIDA, India, 10–12 August 2017.
21. Basari, A.S.H.; Hussin, B.; Ananta, I.G.P.; Zeniarja, J. Opinion mining of movie review using hybrid method of support vector machine and particle swarm optimization. *Procedia Eng.* **2013**, *53*, 453–462. [[CrossRef](#)]
22. Seal, A.; Ganguly, S.; Bhattacharjee, D.; Nasipuri, M.; Gonzalo-Martin, C. Feature Selection using Particle Swarm Optimization for Thermal Face Recognition. In *Applied Computation and Security Systems*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 25–35.
23. Shang, L.; Zhou, Z.; Liu, X. Particle swarm optimization-based feature selection in sentiment classification. *Soft Comput.* **2016**, *20*, 3821–3834. [[CrossRef](#)]
24. Yun, C.; Oh, B.; Yang, J.; Nang, J. Feature subset selection based on bio-inspired algorithms. *J. Inf. Sci. Eng.* **2011**, *27*, 1667–1686.

25. Ranawana, R.; Palade, V. Multi-Classifer Systems: Review and a roadmap for developers. *Int. J. Hybrid Intell. Syst.* **2006**, *3*, 35–61. [[CrossRef](#)]
26. Blitzer, J.; Dredze, M.; Pereira, F. Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. In Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, Prague, Czech Republic, 23–30 June 2007.
27. Porter, M. The Porter Stemming Algorithm. Available online: <http://tartarus.org/martin/PorterStemmer/> (accessed on 20 March 2017).



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).