Research Article GU J Sci 31(3): 789-799 (2018)

Gazi University

Journal of Science

http://dergipark.gov.tr/gujs

An Unsupervised Approach for Selection of Candidate Feature Set Using Filter Based Techniques

Sai Prasad POTHARAJU[1] *, Marriboyina SREEDEVI[2]

[1]*Computer Science and Engineering Department, K L University, Guntur (AP), India, 522502*

[2]*Computer Science and Engineering Department, K L University, Guntur (AP), India, 522502*

Article Info

Received: 14/12/2017
Accepted: 08/05/2018

Keywords

Feature Selection
Clustering
Multi-Layer Perceptron
Min-Max Normalization
Filter Methods

Abstract

High dimensionality is the one of the important issue in preprocessing stage of data mining. Initial feature space may have irrelevant or redundant features. These properties of features decrease the performance of classifier, and also require more memory and high computing power. This issue can be addressed by selecting the best feature subset for improving the classification performance. In this research, we have proposed an unsupervised approach using filter based feature selection methods and K-Means clustering technique to derive the candidate subset. Score of each feature is calculated using traditional filter based methods. Then Min-Max technique is applied to normalize the dataset. K-Means algorithm is employed on the dataset to form the clusters of features. To decide the strong subset, Multi-Layer Perceptron(MLP) is applied on each cluster. Best cluster is selected based on the minimum Root Mean Square (RMS) error rate given by MLP. This framework is compared with traditional methods over six well known datasets having the total features in between 34 and 90 using various classification algorithms. The proposed method recorded 75% competitive rate than Information Gain(IG), 71% success rate than Gain Ratio Attribute Evaluator(GR) and Chi Square Attribute Evaluator(Chi), 83% competitive rate than ReliefF(Rel) traditional methods. Jrip classifier performed 55%, J48 recorded 66%, Naive Bayes displayed 88%, IBK (Instance Based) displayed 80% success rate over all the datasets.

1. INTRODUCTION

High dimensionality can be defined as, having multiple (more number) features in a dataset [1]. Feature Selection (FS) is one of the major issues to be considered in Pre-Processing stage of Data Mining along with class imbalance, noise, missing values, missing labels. There are several problems associated with the high dimensional dataset. Those problems can be addressed using the concept of FS. Selecting the best candidates (features) out of availability of initial dataset is called as FS.
All the features in the primary feature space can't give significant performance always. Because it is always possible that, there might be some noisy, redundant and irrelevant features in the dataset. If these characteristic features are presented in the dataset, it creates several complications to the classifier. Those are: classifier may misbehave in prediction; performance may become lower than expected. Moreover, if all the presented features are taken into account, high computing and memory configuration setup is required for training the model [2].

Research question here is, why we need to consider all the features for classification? Why can't we reduce or select only best features and discard the low significant features?. However, there have been some existing methods like FS and Feature Extraction are available in the literature for addressing those research questions raised. In this study, we mainly focused on FS methods only.

*Corresponding author, e-mail: psaiprasadcse@gmail.com

In the literature, there are three groups of FS methods are existed [3]. Those includes: Filter group, Wrapper, and Hybrid group [4]. The working style of these three groups and their performance are different from each other. Filter group awards the position/rank to each feature as per the feature's worth in classifying an unknown instance. Depending on the kind of the problem or dataset considered, top 'N' features can be selected for classification and remaining can be ignored. Information Gain (IG), Chi-Squared (CHI), Relief-F(Rel), and Gain Ratio(GR) are some of the traditional algorithms belong to filter group[5]. Wrapper approach uses the searching and learning methods. It derives the best candidate subset. This is time consuming process than filter as searching every combination is costly, whereas embedded methods combine the advantages of both these approaches [4]. In this research paper, we tried to articulate a novel feature selection methodology using filter based methods and popular unsupervised clustering technique called K-Means. Rest of the manuscript is articulated in various sections. Some existing related literature is given in second section. Proposed methodology with a simple example is described in section 3. Experiment and dataset description is given in section 4. Result analysis with discussion is produced in section 5. Finally, article is concluded with ending remarks.

2. LITERATURE

Features (Dimensions) 'N' in the scale of few tens to hundreds are generally referred to as high-featured (dimensional) data [1]. FS concepts have employed to tens or hundreds or thousands of features in recent studies by various researchers for different applications [6, 7]. FS is punished by high dimensionality [8]. There are three FS approaches have been proposed for addressing the problems associated with high dimensionality in the existing studies namely: Filter, Wrapper, and Hybrid. FS have been used in many real world applications in recent days. Those applications includes: text categorization, remote sensing, intrusion detection, genomic analysis, image analysis, etc.[9]. Symmetrical Uncertainty based feature selection is proposed by the authors and employed over medical datasets [10]. Authors also proposed feature selection approach based on correlation coefficient to draw the subset of features for improving the classification performance [11].

As our study is based on filter based feature selection algorithms like Information Gain (IG), Gain Ratio (GR), Chi- Squared (Chi), and ReliefF(Rel), in this section some existing theory about those methods are described. IG can be defined as information that is derived by drawing the score of the feature, which is the subtraction of the entropy (E) distribution before the split (dbs) and distribution after the split (das) [12]. The highest IG is equivalent to the lowest E. It can be given by

$$IG=(E(dbs)-E(das)) \tag{1}$$

$$E(x_1,x_2,\dots,x_n)=-x_1log(x_1)-x_2log(x_2)-\cdots-x_nlog(x_n) \tag{2}$$

Where $x_1,x_2..x_n$ are the attributes of the dataset. E(x) is an entropy of an attribute x.

GR is the non-symmetrical measure that is introduced to adjust the bias of the IG [12].It can be given by

$$GR=\frac{IG}{E(X)} \tag{3}$$

Where E (X) is entropy of X

Chi is an alternative method used more frequently [13]. It measures the value of a feature by calculating the value of the chi-squared statistic with respect to the class. The primary hypothesis H_0 is an assumption that the two features are not related each other, and it is evaluated by below formula:

$$\chi2=\sum_{i=1}^{r}*\sum_{j=1}^{c}\frac{(O_{ij}-E_{ij})^2}{E_{ij}} \tag{4}$$

O_{ij}: is frequency of observed.
E_{ij} : is frequency of the expected.
The higher the score of $\chi 2$, the higher the evidence against the hypothesis H_0 is.

Rel attribute evaluation [14], measures the value of a feature by iteratively sampling a record of the data set and accounting the value of the given feature for the nearest record of the same and different class.

Information gain is applied by the authors with various classifiers to study the Indian music data set, they have considered top 11 features and tested with Multi-Layer Perceptron, thereby secured the maximum accuracy[15] . Comparison between IG, GR, Chi, Rel with the some of the supervised algorithms have been presented for ranking the features of Australian and credit approval data sets[16]. IG and Chi also applied for Malay text categorization with Naive Bayes (NB), KNN, and N-gram. Chi squared method recorded 96.14 % accuracy with KNN [17]. FS concept also used in intrusion detection system to increase the prediction rate, as data set contains irrelevant and redundant data, which lower the classification rate[18].

In this current study, we have proposed a technique to form the cluster of features. For clustering purpose the popular K-means clustering is applied. It is an unsupervised learning approach, it can be used when an instance is not associated with a class label and designate the unknown instance cluster in which it can be located.

To determine the optimal number of clusters, Elbow method is applied on each data set. Steps to find the optimal clusters are given below [19].

1. Compute clustering algorithm for various values of k (say 1 to 10).
2. For every k, compute the total $\boldsymbol{\Sigma}$wss (within-cluster sum of square).
3. Plot the curve of wss as per the number of clusters k.
4. The position of a bend (knee) in the plot is normally considered as an indicator of the optimal number of clusters.

As a result of K-means 'K' clusters will be created, each cluster is equipped with unique features. To know the best cluster, MLP have been applied on each cluster which is one of the popular classification techniques.

3. PROPOSED METHODOLOGY

Our ultimate goal of this research is to draw the best candidate subset. For this, initially dataset is balanced using SMOTE (Symmetric Minority over Sampling Technique) to address the class imbalance problem, which is based on K-Nearest Neighbor (KNN) approach [20-22]. In next level, IG, Rel, Chi, GR methods have been employed on the datasets then recorded each attribute's score. As a result of this process, we could form transformed dataset with 4 features, and instances are equal to the number of features in initial dataset. The selected (IG, Rel, Chi, GR) methods are ranking algorithm which are based on the concept of information theory [23]. These methods assign the weight (value) along with rank to each feature, as per it's information worth. There are other feature selection methods like CFS Subset Evaluator, Classifier Subset Evaluator, Consistency Subset Evaluator also existed. Those methods produce the candidate subset based on searching technique and classifier employed. These methods are costly in terms of memory and computation time[24]. Next, transformed dataset is normalized using Min-Max theory of normalization to avoid overfitting[25].Min-Max technique is as follows:

$$z_i = x_i - \min(x) / (\max(x) - \min(x)) \tag{5}$$

where $x=(x_1,...,x_n)$ and z_i is i^{th} normalized data.
The reason to apply Min-Max technique is to normalize the dataset. The data points of IG, Rel, GR are in the range of 0 to 1, but CHI data points are ten to hundreds and hundreds to thousands over some datasets. Distance between data points will be varied while applying the K-means for forming the clusters if it is

not normalized and it may leads to overfitting.
Normalized dataset is now unsupervised, it means there is no class label in it. As our transformed dataset does not have any class label, K-means technique is applied. The purpose of applying the K-means here is to find out the similar points (feature) which can share common properties. By applying this method all strong features can be in one clusters, all weak features can be in another cluster. Next, normalized dataset is divided into number of clusters using K-Means technique. For optimal number of clusters, Elbow method is employed. After the result of clustering, similar features are grouped in same cluster. We don't require all the clusters. We need only best or strong cluster. For this, MLP is applied on each cluster then depending on the minimum RMS error rate strong cluster is selected. Features in strong cluster are considered as best candidate subset. The same procedure is presented in algorithmic way.

Algorithm

- Check the class imbalance, then apply the SMOTE if not balanced.
- Apply IG, Rel, Chi, GR on balanced dataset and record each attribute's score.
- Normalize the attribute's score using Min-Max technique and get the normalized dataset.
- Define the optimal number of clusters (K) by applying Elbow method over normalized dataset.
- Apply K-Means algorithm and find the clusters.
- Apply MLP on K clusters.
- Choose the best cluster based on the RMS error rate.
- Apply the classification techniques with the features in the best cluster.

The proposed algorithm is represented as flowchart with Figure 1.

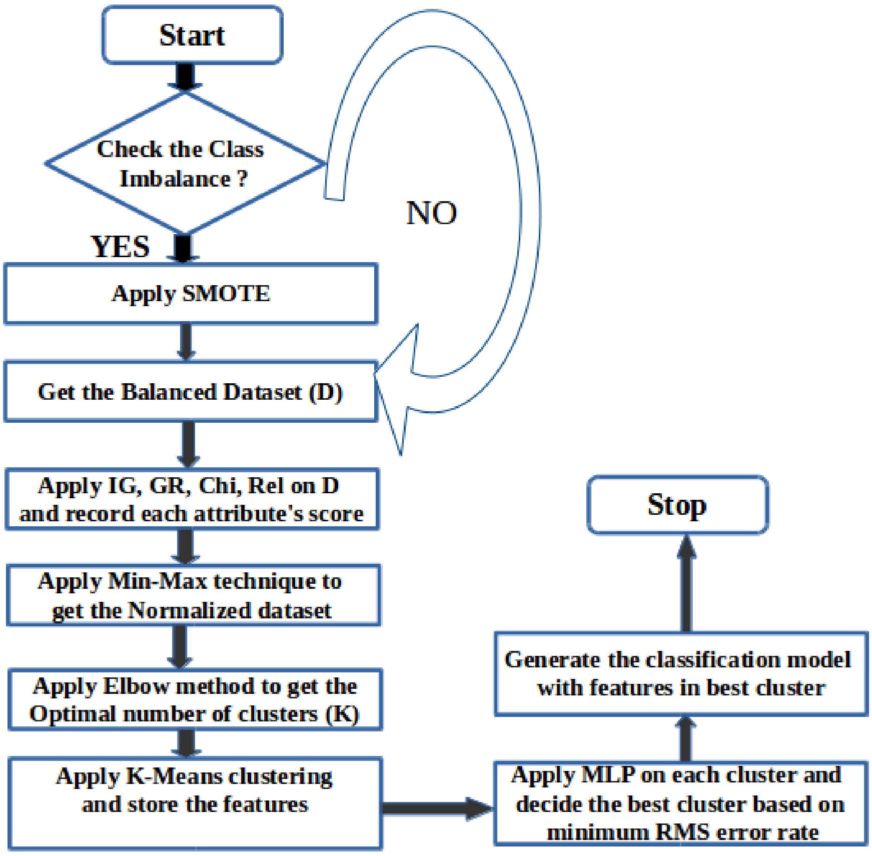

Figure 1. *Proposed methodology*

Example

Assume there are 10 features in the initial balanced dataset. Their score by filter based methods IG, Rel, Chi, GR is like in Table 1.

Table 1. *Sample attribute's score by filter based method*

FNO	IG	Rel	Chi	GR
1	0.41	0.20	216	0.35
2	0.29	0.195	214	0.29
3	0.28	0.16	190	0.27
4	0.27	0.15	180	0.26
5	0.26	0.14	150	0.25
6	0.25	0.13	120	0.24
7	0.24	0.13	110	0.22
8	0.23	0.12	105	0.21
9	0.22	0.11	103	0.19
10	0.21	0.09	99	0.17

From the above table, it is clear that all the values are not in the same range, so it needs to be normalized. For this, apply the Min-Max technique over Table 1 dataset, normalized dataset can be available in Table 2 after applying Min-Max technique.

Table 2. *Normalized dataset*

FNO	IG	Rel	Chi	GR
1	1	1	1	1
2	0.4	0.95	0.98	0.67
3	0.35	0.64	0.78	0.56
4	0.30	0.55	0.69	0.50
5	0.25	0.45	0.44	0.44
6	0.20	0.36	0.18	0.39
7	0.15	0.36	0.09	0.28
8	0.10	0.27	0.05	0.22
9	0.05	0.18	0.03	0.11
10	0.00	0.00	0.00	0.00

Apply Elbow method over Table 2 dataset to find the optimal number of clusters (K), assume Elbow method is recommended K=3. Then , form 3 clusters by applying K -Means over unsupervised dataset (Table 2).After this step, apply the MLP on each cluster with the features in it and record its RMS. Based on its result decide the best cluster. Table 3 shows the features formed in cluster and RMS error rate derived by MLP.

Table 3 *.Cluster of features and RMS error rate*

Cluster Id (Cid)	Features in it	RMS
C1	1, 2, 3	0.409
C2	7, 8, 9	0.423
C3	4, 5, 6	0.315

Cluster 3 (C3) has minimum error rate, so C3 can be designated as a strong cluster and features in it (4, 5, 6) can be designated as best candidate subset.

4. EXPERIMENT

The proposed framework is evaluated on six real time well known datasets available in public domain [26]. Their description is given below in Table 4. The experiment is demonstrated using WEKA machine learning tool with all default settings for training and testing the classification accuracy [27]. K-Means algorithm is applied to form the clusters. Elbow method is applied over the dataset to decide the optimal number of clusters. It is implemented using R statistical programming language.

The sample code to know the optimal number of clusters is given below.

```
library (NbClust)
df <- read.csv("dataset.csv")
nb <- NbClust(df, distance = "euclidean", min.nc = 2,max.nc = 10, method = "kmeans")
```

In the above function euclidean is used for measuring the distance between feature to centroid of cluster and cluster to cluster, min.nc is the value of minimum number of clusters, max.nc is the maximum number of clusters.

To compare the knowledge of proposed approach, top ‘N’ features derived by filter based algorithms are taken into account. Where ‘N’ is equal to the number of features in the strong cluster. For a moment, in the given example C3 is the strong cluster and it has 3 features in it, so top 3 features derived by filter methods are considered. To know the significance of the proposed method 4 different types of classification algorithms are applied namely Jrip, J48, Naive Bayes (NB), K-Nearest Neighbor (IBK).

Table 4*. Dataset description*

Dataset ID	Dataset Name	#Features
D1	Bio Degradation	41
D2	Sonar	60
D3	Spambase	57
D4	Movement Libras	90
D5	Ionosphere	34
D6	KDD (Intrusion Detection System)	40

5. RESULTS AND DISCUSSION

This section comprises the results originated by the candidate subset derived by proposed methods over the various datasets using different classification algorithms with possible discussion. Number of clusters(K), number of features(F) derived by proposed method and RMS error rate after applying MLP on each cluster is produced left side of each table, classification performance with best candidate subset and top features derived by traditional methods using various classifiers is produced right side of each table. Result over Bio Degradation dataset is given in Table 5.

Table 5. *RMS error rate on each cluster and classification performance (Bio Degradation)*

RMS error rate after applying MLP on each cluster				Classification performance with best candidate subset and top features derived by traditional methods				
K	Cid	F	RMS		Jrip	J48	NB	IBK
5	C1	10	.4143	**C2**	**84.47**	**84.83**	**77.61**	**83.47**
	C2	**11**	**.3559**	IG	<u>83.40</u>	<u>83.90</u>	<u>77.61</u>	<u>83.47</u>
	C3	11	.4521	GR	<u>81.83</u>	85.05	<u>77.53</u>	<u>82.68</u>
	C4	1	.4997	CHI	<u>83.54</u>	<u>84.12</u>	78.82	83.54
	C5	8	.3731	REL	<u>84.40</u>	85.55	<u>77.46</u>	84.12

Note: K : Optimal number of clusters derived by Elbow method.
Cid: Cluster ID.
F: Number of features in each cluster.
RMS: Root Mean Square error rate

Over Bio Degradation dataset, Elbow method has recommended 5 clusters (K), C2 cluster has 11 features and its RMS error rate is minimum, so it is designated as best cluster, and features in it are derived as best candidate subset. Initially there are 41 features in it, after applying the proposed method it has produced 11 strong features. The subset derived by proposed method performed better than all traditional methods using Jrip. Also C2, performed better than existing IG, CHI using J48. C2 Recorded little improvement than existing GR and Rel, also competing with IG using NB. It also displayed competitive performance with IG and GR using IBK. Result over Sonar dataset is given in Table 6. Out of 60 features, 18 features have been derived.

Table 6. *RMS error rate on each cluster and classification performance (Sonar)*

RMS error rate after applying MLP on each cluster				Classification performance with best candidate subset and top features derived by traditional methods				
K	Cid	F	RMS		Jrip	J48	NB	IBK
4	C1	35	.4092	**C2**	**71.10**	**76.60**	**72.93**	**80.27**
	C2	**18**	**.4076**	IG	81.19	<u>76.60</u>	<u>69.26</u>	86.23
	C3	5	.4509	GR	<u>70.18</u>	<u>75.68</u>	<u>68.34</u>	<u>79.81</u>
	C4	2	.4141	CHI	77.98	78.44	<u>68.34</u>	85.32
				REL	77.98	77.52	<u>67.88</u>	88.53

Elbow method has recommended 4 clusters (K), C2 cluster has 18 features and its RMS error rate is minimum, so it is nominated as the best cluster, and features in it are derived as best candidate subset. The subset derived by proposed method performed better than traditional GR methods using Jrip. Also C2, competing with existing IG, GR using J48. C2 recorded better than all traditional methods using NB.

It also displayed competitive performance than GR using IBK. Result over Spambase dataset is given in Table 7. Out of 57 features, 19 features have been derived.

Table 7. *RMS error rate on each cluster and classification performance (Spambase)*

RMS error rate after applying MLP on each cluster				Classification performance with best candidate subset and top features derived by traditional methods				
K	Cid	F	RMS		Jrip	J48	NB	IBK
4	C1	38	.3675	**C2**	**92.67**	**93.48**	**88.70**	**90.27**
	C2	**19**	**.2559**	IG	92.67	93.48	88.70	90.27
				GR	90.52	91.66	76.43	89.57
				CHI	92.67	93.48	88.70	90.27
				REL	90.22	90.90	74.20	89.01

Elbow method has recommended 2 clusters (K), C2 cluster has 19 features and its RMS error rate is minimum, so it is nominated as the best cluster, and features in it are derived as best candidate subset. Proposed method over Spambase dataset performed better than all other existing filter methods.
Result over Movement Libras dataset is given in Table 8. Out of 90 features, 32 features have been derived.

Table 8. *RMS error rate on each cluster and classification performance (Movement Libras)*

RMS error rate after applying MLP on each cluster				Classification performance with best candidate subset and top features derived by traditional methods				
K	Cid	F	RMS		Jrip	J48	NB	IBK
4	C1	16	.2253	**C3**	**46.66**	**64.72**	**55**	**82.22**
	C2	26	.16	IG	48.05	64.16	51.94	76.94
	C3	**32**	**.15**	GR	49.16	60.27	47.77	75
	C4	16	.20	CHI	47.5	62.22	49.16	77.77
				REL	45	60.55	51.66	78.83

Elbow method has recommended 4 clusters (K), C3 cluster has 32 features and its RMS error rate is minimum, so it is considered as the best cluster, and features in it are derived as best candidate subset. Over Movement Libras dataset, features drawn by proposed method have performed better than all other existing filter methods using J48, NB and IBK. Result over Ionosphere dataset is given in Table 9. Out of 34 features, 23 features have been derived.

Table 9. *RMS error rate on each cluster and classification performance (Ionosphere)*

RMS error rate after applying MLP on each cluster	Classification performance with best candidate subset and top features derived by traditional methods

K	Cid	F	RMS		Jrip	J48	NB	IBK
2	**C1**	**23**	**.2302**	**C1**	**87.06**	**89.43**	**86.85**	**94.61**
	C2	11	.3503	IG	87.06	89.43	86.85	94.61
				GR	87.28	89.87	87.28	93.96
				CHI	87.06	89.43	86.85	94.61
				REL	85.99	89.87	86.42	94.18

Elbow method has recommended 2 clusters (K), C1 cluster has 23 features and its RMS error rate is minimum, so it is considered as the best cluster, and features in it are derived as best candidate subset. Over Ionosphere dataset, features drawn by proposed method produced competitive result than IG, CHI, REL with Jrip, J48, NB. It is also performed better than all traditional methods with IBK. Result over KDD dataset is given in Table 10. Out of 40 features, 15 features have been derived.

Table 10. *RMS error rate on each cluster and classification performance (KDD)*

RMS error rate after applying MLP on each cluster				Classification performance with best candidate subset and top features derived by traditional methods				
K	Cid	F	RMS		Jrip	J48	NB	IBK
2	C1	25	.147	**C2**	99.48	**99.59**	**91.58**	**98.86**
	C2	**15**	**.135**	IG	99.57	99.64	89.21	98.41
				GR	99.51	99.54	89.11	98.76
				CHI	99.57	99.64	89.21	98.41
				REL	99.51	99.11	89.85	99.10

Elbow method has recommended 2 clusters (K), C2 cluster has 15 features and its RMS error rate is minimum, so it is nominated as the best cluster, and features in it are derived as best candidate subset. Over KDD dataset, features drawn by proposed method produced competitive result than GR and REL with J48. Also the proposed method displayed better results than all traditional approaches with NB. It is also observed that proposed method competing with traditional IG, GR, CHI with IBK.

Proposed method produced various results depending on the dataset. Most of the cases it is recorded competitive results than traditional methods. Because each cluster is built with different features. For example, as per the outcome of proposed methodology over Bio Degradation dataset, C2 is the best cluster and it has 11 features in it. Those feature ids are: 5, 9, 11, 12, 13, 15, 18, 23, 34, 38, 41. Top 11 features derived by existing methods are different from the features derived by proposed method. Because of this reason over some dataset result is high and some cases result is low.

The proposed method recorded 75% competitive rate than Information Gain(IG), 71% success rate than Gain Ratio Attribute Evaluator(GR) and Chi Square Attribute Evaluator(Chi), 83% competitive rate than ReliefF(Rel) shown competitive performance than few of the traditional methods. Jrip classifier performed 55%, J48 recorded 66%, Naive Bayes displayed 88%, IBK displayed 80% success rate over all the datasets.

6. CONCLUSION

In this research work we have presented an unsupervised feature selection approach using filter based ranking algorithms for addressing the high dimensional issue of preprocessing. For the proposed work, we initially considered balanced dataset and applied 4 basic filter based ranking methods. Attribute's score is calculated using those filter methods. After this process, dataset is divided into number of optimal clusters. Elbow method is applied to decide the optimal number of clusters. Each such cluster has reduced number of features in it. One strong cluster is selected by applying MLP on each cluster which is formed by proposed method. Based on the minimum RMS error rate given by MLP the best cluster is decided. This approach is tested with 4 different classifiers over six datasets then compared with traditional methods. As a result of this method, approximately 35-40% (Average of all) of features can be selected for classification, thereby memory consumption can be reduced and performance can be increased.

CONFLICTS OF INTEREST

No conflict of interest was declared by the authors.

REFERENCES

[1] Liu, H., Yu, L., "Toward integrating feature selection algorithms for classification and clustering", IEEE Transactions on knowledge and data engineering, 17(4): 491-502,(2005).

[2] Liu, C., Wang, W., Zhao, Q., Shen, X., Konan, M.," A new feature selection method based on a validity index of feature subset", Pattern Recognition Letters, 92(1):1-8,(2017).

[3] Stańczyk, U., "Feature evaluation by filter, wrapper, and embedded approaches", In Feature Selection for Data and Pattern Recognition,584(1): 29-44,(2015).

[4] Chandrashekar, G.,Sahin, F.,"A survey on feature selection methods", Computers & Electrical Engineering, 40(1): 16-28,(2014).

[5] 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 Asia information retrieval symposium,8870(1):429-443,(2014).

[6] Kumar, V., Minz, S.,"Poem classification using machine learning approach", In Proceedings of the Second International Conference on Soft Computing for Problem Solving,236(1): 675-682,(2012).

[7] Kumar, V., Minz, S., "Multi-view ensemble learning for poem data classification using SentiWordNet", In Advanced Computing, Networking and Informatics,1(1): 57-66,(2014).

[8] Friedman, J., Hastie, T.,Tibshirani, R., "The elements of statistical learning ",Springer series in statistics,1(1):337-387,(2001).

[9] Kumar, N., "Literature Review-Chronic Regulatory Focus and Financial Decision-Making". Springer Briefs in Finance,1(1):5-20,(2016).

[10] Potharaju, S. P., Sreedevi, M., "A Novel M-Cluster of Feature Selection Approach Based on Symmetrical Uncertainty for Increasing Classification Accuracy of Medical Datasets", Journal of Engineering Science & Technology Review, 10(6):154-162,(2017).

[11] Potharaju, S. P., Sreedevi, M., "A Novel Clustering Based Candidate Feature Selection Framework Using Correlation Coefficient for Improving Classification Performance", Journal of Engineering Science and Technology Review, 10(6): 38-43,(2017).

[12] Hall, M. A., Smith, L. A.,"Practical feature subset selection for machine learning", Proceedings of the 21st Australasian Computer Science Conference ACSC, 1(1):181-191,(1998).

[13] Liu, H., Setiono, R., "Chi2: Feature selection and discretization of numeric attributes", Proc. IEEE 7th International Conference on Tools with Artificial Intelligence, 1(1): 338-391,(1995).

[14] Robnik-Šikonja, M., Kononenko, I. ,"Theoretical and empirical analysis of ReliefF and RReliefF", Machine learning, 53(1): 23-69,(2003).

[15] Kalapatapu, P., Goli, S., Arthum, P., Malapati, A.,"A Study on Feature Selection and Classification Techniques of Indian Music", Procedia Computer Science, 98(1): 125-131,(2016).

[16] Novaković, J.,"Toward optimal feature selection using ranking methods and classification algorithms", Yugoslav Journal of Operations Research, 21(1):119-135,(2016).

[17] Alshalabi, H., Tiun, S., Omar, N., Albared, M. , "Experiments on the use of feature selection and machine learning methods in automatic malay text categorization", Procedia Technology, 11(1): 748-754,(2013).

[18] Hasan, M. A. M., Nasser, M., Ahmad, S., Molla, K. I.,"Feature selection for intrusion detection using random forest", Journal of information security, 7(03): 129-140,(2016).

[19] Charrad, M., Ghazzali, N., Boiteau, V., Niknafs, A. ,"NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set", Journal of Statistical Software, 61(6): 1-36,(2014).

[20] Chawla, N. V., Bowyer, K. W., Hall, L. O., Kegelmeyer, W. P. ,"SMOTE: synthetic minority over-sampling technique", Journal of artificial intelligence research, 16(1): 321-357,(2002).

[21] Potharaju, S. P., Sreedevi, M., "An Improved prediction of Kidney disease using SMOTE." ,Indian Journal of Science and Technology ,9(31):1-7, (2016).

[22] Potharaju, S. P., Sreedevi, M., "Ensembled Rule Based Classification Algorithms for predicting Imbalanced Kidney Disease Data" ,Journal of Engineering Science and Technology Review ,9(5): 201-207,(2016).

[23] Singh, B., Kushwaha, N., Vyas, O. P.,"A feature subset selection technique for high dimensional data using symmetric uncertainty", Journal of Data Analysis and Information Processing, 2(04), 95-105,(2014).

[24] Kumar, C. A., Sooraj, M. P.,Ramakrishnan, S.," A Comparative Performance Evaluation of Supervised Feature Selection Algorithms on Microarray Datasets". Procedia Computer Science, 115(1): 209-217,(2017).

[25] Patro, S., Sahu, K. K.,"Normalization: A preprocessing stage", Cornell University Library,15(03):64-62 ,(2015).

[26] Dua, D., Karra Taniskidou, E.," UCI Machine Learning Repository",School of Information and Computer Science,Irvine, CA: University of California,1(1):1,(2017).

[27] Witten, I. H., Frank, E., Hall, M. A., Pal, C. J. ," The WEKA Workbench Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufmann, 4(1):285-334,(2016).