2022 International Conference on Emerging Smart Computing and Informatics (ESCI)
AISSMS Institute of Information Technology, Pune, India. Mar 9-11, 2022

Structured Ranking Method-based Feature Selection in Data Mining

Hemanta Kumar Bhuyan[1], Biswajit Brahma[2], S.Nyamathulla[3], Srikanta Kumar Mohapatra[4]
[1,3]Department of IT, Vignan's Foundation for Science, Technology & Research, Deemed to be university, Guntur, Andhra Pradesh, India
[2]Mckesson Corporation, San Francisco CA, USA
[4]Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India
[1]hmb.bhuyan@gmail.com* , [2]Biswajit.Brahma@gmail.com, [3]nyamath.j@gmail.com. [4]srikanta.mohapatra@chitkara.edu.in

2022 International Conference on Emerging Smart Computing and Informatics (ESCI) | 978-1-6654-0073-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/ESCI53509.2022.9758354

***Abstract*—Feature selection has been emphasized on an operative approach for dealing with large volume data. The majority of these approaches are skewed into high-ranking features to get well right features towards classification. This paper proposes a structured feature ranking (SFR) approach for large volume data to address this challenge. We present a subspace feature-based clustering approach to find out feature-based cluster as per class labels. The various feature clusters are created ranked for features independently using the SFR approach, based on the subspace weight provided by SFC. Then, for ranking the features, we offer a structured feature weighting method in which the high-rank characteristics are utilized for class labels. SFC's approach has been tested in a variety of features. On a collection of large volume datasets, the proposed SFR approach is compared to six feature selection methods. The results demonstrate that SFR method outperformed than methods.**

Keywords— Data mining, multi-objective optimization, clustering algorithms, feature selection

I. INTRODUCTION

The large volume data constitute a significant obstacle to unsupervised learning [1] in different machine learning applications through discriminative projection for feature selection. Many learning models overfit and become unintelligible as they classify such data since only a tiny percentage of genes are strongly associated with samples. In contrast, the vast majority of genes are irrelevant. It can solve this challenge using feature selection, which selects the ideal feature set from high-dimension data that contains discriminative characteristics. To deal with high-dimensional data, feature selection has become increasingly significant over the years [2,3,4,5]. Feature ranking is a prominent technique for choosing features with order from feature set. Basically, feature selection methods are divided into filter, wrapper, and embedding methods to evaluate many data sets. Data intrinsic qualities are used to pick feature subsets in the filter techniques, which do not involve any learning process. Fisher score [6] and norm quality [7,8] are standard unsupervised filter algorithms. To evaluate a feature subset, wrapper methods utilize the objective approaches for prediction. It is possible to acquire strong prediction performance using embedded approaches based on training. Embedded approaches outperform the other two by a wide margin and are therefore more prevalent [9,10,11,12,13]. The most common criteria for evaluating the degree of dependency are correlation measurements between the feature and the class.

As a result, the methods mentioned above work best for statistically independent qualities but fall short in recognizing set of features that can be used to forecast class membership. Their focus is on high-ranking characteristics, yet these features may be highly connected. Because linked characteristics may have similar attributes and be redundant, we want to use the fewest correlations possible for classification tasks. We provide a new method called structured Feature Ranking (SFR) to select supervised features from large, high-dimensional datasets. First, we present a subspace feature clustering (SFC) approach to discover feature-based clusters as per feature value related class. The SFC consumes the class labels to extend the Subspace Weighting Co-Clustering (SWCC) [14]. When SFC's co-clustering results are used, each feature cluster's subspace weights are used to rank the features within it. As per closed feature cluster associated with others, we suggest a structured feature weighting approach to determine the highest-ranking features, which are also informative and diverse. To find out how well our methods worked, we ran tests on both fictitious and real-world data. A total of 5 large-scale datasets, comprising 5 gene expression data sets and 7 image datasets are used to compare SFR with other feature ranking algorithms. SFR beat the other feature ranking algorithms on the majority of results, according to the results. The coordination between performance and SFR parameters is also looked over the data set. Our strategy selects features that are very much effective as evidenced by the performance of experiments. As a result, SFR works well with large datasets.

The remaining part of this paper is arranged as follows. We considered the relative work of this paper in section II. In section III, we developed the proposed methodology related to feature selection approaches. The experimental results are well analyzed in section IV. We concluded the paper in section V.

II. BACKGROUND

Feature selection and co-clustering are discussed briefly in this section. In the context of predictive modelling, this process is known as feature selection or variable selection. In the early stages of feature selection approach, linear regression is the most commonly used method. Classification and clustering problems have been added to this approach gradually. There have been numerous approaches to feature selection over the past few decades. There has been a lot of approaches for selecting features from feature set eliminate unnecessary features and improve performance. Filter approach, wrapper approach, and embedded approach are the three main categories of feature selection methods. There is no learning algorithm involved in the filter methods, which choice feature subsets based on the data's essential appearances. Norm quality [7,8] is one of the most common supervised filtering methods [9, 21]. Such methods can be time-consuming because they treat forecasting the performance with the help of different objective components [15]. The training process for embedded methods includes feature selection. It has become increasingly popular to use embedded methods, whose performance is better than others [10,11].

On the other hand, these methods tend to focus on the most highly ranked features, even though these features may be highly correlated. Max-Relevance and Min-Redundancy were used by Peng et al. [16] to select the most relevant features. The first-order incremental approach was used to achieve the best possible features. The correlation methods were incorporated into the support vector machine and recursive feature elimination process by Yan et al. [17]. Some of the reasons given by Das et al. [18] included the following: it can be sure that the selected features are not superfluous and are more demonstrative of the unique feature space because of the correlations between them. For classification tasks, it focused on features that are more distinct and have the fewest correlations, since these features may have similar properties. It suggested a wrapper approach to select feature that attempts to forecast class labels on less features using linear regression. However, it takes a long time to implement. An uncorrelated feature selection was proposed by Kong et al. [19]. A two-feature group will be formed as per manual threshold of correlation. Each feature group will be given the standard $l_{2,1}$ regularization to reduce the number of highly correlated feature pairs and choice the most significant features in the majority of feature sets. However, determining a proper threshold is difficult. For example, the quantity of feature set may be overlarge, obscuring some of the most important ones. As a result, most correlations will be ignored because there aren't enough feature groups to cover all of them.

The Co-Clustering is developed by using both records and attributes of a database [20], or bi-clustering. Text mining, bioinformatics, and recommendation systems are just a few of the areas where it has recently been used. Co-clustering methods are proposed as an alternative to existing clustering methods that help to analyze the features as per feature ranking, To put it another way, it is considered the clusters with distributed manner of sharing of data. It is common practice to use co-clustering methods to analyses the hidden construction that involves set of features and to improve the clustering performance in these kinds of data sets. It has been suggested that hierarchical and several clustering such as spectral and partitional co-clustering [2] are the best models for co-clustering.

III. FRAMEWORK OF STRUCTURAL FEATURE RANKING METHOD

We have considered the structural feature ranking approaches for choosing feature. We developed cluster with high correlated features and choose features from feature set with the help of cluster to minimize the redundancy. Chen et al. [14,22] had made earlier cluster based on high correlated features in the form of subspace. We present a structured ranking method for feature selection in this paper. The new method begins by grouping the features into a set of feature clusters, which we'll be considered for feature ranking as per method. A structured weighting ranking method make list of features as their ranking as per SFR and feature clusters which is proposed to obtain a concluding list of features as per ranked from various feature clusters.

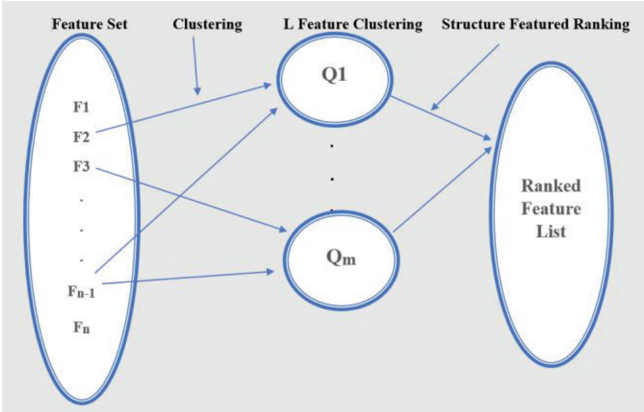

Fig. 1. Process of structured featured ranking method

We proposed the structured feature ranking method's steps for performing the analysis of selected feature as shown in fig 1. To find the disjoint feature clusters, we first group the labelled dataset X with n number of features as $F = \{f_1,\dots f_n\}$ into number of disjoint feature clusters as $\{Q_1 \dots Q_m\}$, where $Q_j \cap Q_i = \Phi$ ($\forall$ i≠ j) and $\bigcup_{j=1}^{m} Q_j = F$ Finally, we use a structured weighting feature ranking method to determine the order in which the n features should be ranked. We'll go over feature clustering and structured weighting in these sections as follow.

A. Optimized Features

Let the labeled data matrix $X \in R^{nxm}$ contain n instances and m features. We follow a subspace weight matrix (SWM) $C \in R^{kxl}$ as [14] with c_{gj} weight based on j^{th} column and g^{th} row cluster. But we considered labelled data matrix with k-row and l-column for cluster X. We also developed the objective function based on Subspace weighting co-clustering (SWCC) as [14] in eq. 1.

$$\min_{U,V,Z,C} \frac{1}{mn} \sum_{g=1}^{k} \sum_{h=1}^{l} \sum_{i=1}^{n} \sum_{j=1}^{m} u_{ig} v_{jh}\, c_{gj} (x_{i,j} - z_{g,h})^2 \quad (1)$$

$$+ \frac{\eta}{m} \sum_{g=1}^{k} \sum_{j=1}^{m} c_{gi} log c_{gi}$$

$$\text{s.t.} \quad \sum_{g=1}^{k} u_{ig} = 1,\; u_{ig} \in \{0,1\},$$

$$\sum_{h=1}^{l} v_{jh} = 1,\; v_{jh} \in \{0,1\},$$

$$\sum_{j=1}^{m} c_{gj} = 1,\; c_{gj} \in \{0,1\}$$

Base on class label in X , it is possible to construct $U \in R^{nxk}$ with known class labels by setting $u_{ig} = 1$ when $x_i \in g^{th}$ class and 0, otherwise. This is called supervised feature selection. The feature clustering process's goal is to group n features in cluster X and create l number of feature clusters. Getting this goal, we considered SFC to generate the objective function as eq. 2 from eq. 1.

$$\min_{V,Z,C} \frac{1}{mn} \sum_{g=1}^{k} \sum_{h=1}^{l} \sum_{i=1}^{n} \sum_{j=1}^{m} u_{ig} v_{jh}\, c_{gj} (x_{i,j} - z_{g,h})^2$$
$$+ \frac{\eta}{m} \sum_{g=1}^{k} \sum_{j=1}^{m} c_{gi} log c_{gi} \qquad (2)$$

s.t. $\sum_{h=1}^{l} v_{jh} = 1,\ v_{jh} \in \{0,1\},$

$\sum_{j=1}^{m} c_{gj} = 1,\ c_{gj} \in \{0,1\}$

The eq. (2) has the approximately same solution as eq. (1) of V, Z and C. As per [4], we can find the solution of V, Z and C from eq. 2 which are explained as below. When Z and C are static, then V can be solved as follows,

$$\begin{cases} v_{jh} = 1, if\ P_{(h)} \le P_{(t)}\ for\ 1 \le t \le L\ where \\ P_{(t)} = \sum_{g=1}^{k} \sum_{i=1}^{n} u_{ig} c_{gj} (x_{ij} - z_{gt})^2 \\ v_{jt} = 0\ for\ t \neq h \end{cases} \qquad (3)$$

The optimal solution for Z is defined as fixed value of v and c as eq. (4).

$$z_{gh} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{m} u_{ig} v_{jh} c_{gj} x_{ij}}{\sum_{i=1}^{n} \sum_{j=1}^{m} u_{ig} v_{jh} c_{gj}} \qquad (4)$$

Similarly, the optimal solution for C is determined by fixed value of both Z and V as eq. (5)

$$c_{gj} = \frac{\exp\{-\frac{E_{gj}}{\eta}\}}{\sum_{j'=1}^{m} \exp\{-\frac{E_{gj'}}{\eta}\}} \qquad (5)$$

Where

$$E_{gi} = \frac{1}{n} \sum_{h=1}^{l} \sum_{i=1}^{n} u_{ig} v_{jh} (x_{ij} - z_{gt})^2 \qquad (6)$$

Based on above equations, we developed the algorithm 1 for the work flow of feature clustering ranking based on subspace weighted matrix

……………………………………………………..
Algorithm 1: Subspace based Feature Clustering
………………………………………………………
1: Input: Dataset X, l- feature clusters and parameter η.
2: Output: SFC result V and the SWM C.
3: Make a matrix $U \in R^{n \times k}$
where $u_{ig} = 1$ for the i^{th} feature with the g^{th} class.
4: initiate i from 0
5: start from Z and $c_{gj} = 1/m$ for ∀ g and j.
6: repeat
7: Compute V^{i+1} using (3).
8: Compute Z^{i+1} using (4).
9: Compute C^{i+1} using (5) and (6).
10: increment of i by 1
11: till getting SFC result V and the SWM C with local minimum value
………………………………………………………………

The Algorithm 1 precis the comprehensive information to the eq. (2) with updating V, Z, and C in a cyclic fashion until convergence is achieved. Because we're getting closer and closer to the local minima of the equ. (2). The optimization process is strictly decreasing to local minima. SFC's computational complexity is O (rnmkl), when the algorithm converges after r iterations. Using k-means, we can see that clustering large high-dimensional data can be efficient because the computational cost of SFC is proportional to the number of features and the number of records. The SFC algorithm is searching center of clusters to initiate and generate several feature clusters with dissimilar starting clusters. Then we run SFC with different initial cluster centers for each of l and η to provide a co-clustering output Ĥ. Each object's label is predicted by applying the learned V^*, Z^*, and C^* to determine the class in which it belongs. This is done by placing the object in the class where it has the least weighted distance from the others.

$$label\ (x_i) = \arg\min_g [\sum_{h=1}^{l} \sum_{j=1}^{m} v_{jh}^* \, c_{gj}^* \, (x_{ij} - z_{gh}^*)^2] \qquad (7)$$

After that, various evaluation indices, such as accuracy, recall, and others, can be used to assess the accuracy of the classification output getting from H. The user typically specifies the number of feature clusters l to use. The best co-clustering result can be chosen from a set of multiple co-clustering results by selecting multiple l and dissimilar clusters $Q_1,..., Q_l$ to rank features.

B. Structured weighting feature ranking

Since each feature's contribution to each class is identified by C ϵ H*, accordingly ranking the features to C is logical. As a result, a projection matrix W∈ R^{mxk} is learned, and the importance of the features can be estimated as $\{\|w^1\|_2 \ldots \|w^m\|_2\}$ using the least-square regression method. Due to the non-negative nature of the C in SFC, we can assess the relative importance of various features using the formulas $\{\|c_1\|_1 \ldots \|c_m\|_1\}$. The nominated features are involved in a few feature clusters with well correlation if it choose r high-rank features as C which can select r import features. We proposed a ranked weighting method for ranking features to choose features with well performance. The new method sorts feature in feature clusters according to $\{\|c_1\|_1 \ldots \|c_m\|_1\}$, then it uses in ascending order to sort the remaining features in the clusters in reverse order. Considering that the j-th feature index for cluster as l_j, we execute a weighted feature ranking vector θ ∈ R^m with a structured weighting as θ_j is defined as

$$\theta_j = \|c_j\|_1 \lambda^{l_j} \qquad (8)$$

where the user specifies the weighting parameter λ ∈ (0,1]. In this case, the weights in a feature cluster are geometrically decreased using λ^{l_j}. To degenerate into the traditional ranking approach, set λ = 1 and use θ_j. If λ<1, it will be geometrically decreased weights, de-emphasizing features with lower order. As a result, selecting various features from a feature cluster will be avoided. Consequently, using's criteria, we can choose features from cluster as per θ.

C. High-Performance Ranking

The algorithm 2 (Structured Feature Ranking) summarizes the above method's detailed procedure. First, we use SFC to create l disjoint feature clusters from m features in X using the new method. Finally, we use a structured weighting feature ranking method to determine how the m features should be ranked. We also modified the algorithm of [24] and use in our proposed model. Readers can refer [24] for details.

………………………………………………..

Algorithm 2: Structured Feature Ranking (SFR)

………………………………………………..

1: Input: Dataset X, l- feature clusters, η, λ are parameters and repeated number of clustering rep.

2: Output: Ranked feature

3: Start with clustering output list $\boldsymbol{H}$ =0.

4: do

5: Execute SFC(X, l, η) and start cluster centers H with results

6: Combine H into $\boldsymbol{H}$.

7: while (j < rep)

8: Authenticate the output H $\in \boldsymbol{H}$, and choose $H^* \in R$ for good performance on clustering

9: Evaluate the $\{||c_1||_1 \ldots ||c_m||_1\}$ and sorting features cluster

10: Evaluate $\theta \in R^{m\times 1}$ as Eq. (8).

11: Arrange top to bottom as top r ranked features with the help of θ.

……………………………………………….

IV. EXPERIMENTAL ANALYSIS

In this section, we considered the 5 gene expression datasets as table 1 for experiments. Different gene express data with number of genes, number of patients and classes are mentioned in table 1. We execute different experiments on dataset as table 1. It shows off the system's capabilities and investigate how it ranks features based on their performance of SFR. As per experimental set up, we considered the dataset D1 as a 100-row, 100-column. D1 is capable of being divided into 16 equal blocks. D1 was used to investigate the SFC algorithm's subspace weights in the experiments. Because the data contains four co-clusters, we chose L = 4 and 20 real values for η. Because initial clusters impact final clustering results, we generated 100 initial cluster centres at random and compared the results. To sum it up, we gathered various output on Structured Feature Clustering for final co-clustering result.

TABLE I. 5 GENE EXPRESSION DATA SETS

Name	Abbr.	No. of Genes	No. of Patients	No. of Classes
Beast.3.class	BR3	4.869	96	3
SRBCT	ST	2.308	83	4
Brain-tumor2	BT2	10.367	50	4
11-tumors	11T	12.533	174	11
14-tumors	14T	15,009	308	26

A. Use of η on C

As per clustering output, we calculated the average entropy of C. When η is small, the average entropy of C decreases. It grew when multiplied by η, then shrank back down. As a result, it grew rapidly as η increase. Entropy regularizes forces weights to be more evenly distributed, so the overall average entropy of C doesn't update more when it's high.

B. Effects of η on the results of feature clustering.

All feature clustering results are evaluated using the five most widely used evaluation indices. Considering that the clustering result depended on initial clusters, we averaged out 100 evaluations and presented the average results. When η is small, everything is low, and then they quickly increase. The results of confusion evaluation based on η parameter is shown in table 2, where the parameter values are given in first column and the corresponding evaluation items such as Accuracy, Precision, Recall, F-measure are mentioned in table -2.

TABLE II. FEATURE CLUSTERING OUTPUTS OF SFC VERSUS H ON D_1

$\eta\downarrow$	Acc	Prec	Recall	f-mes
2^{-17}	0.925	0.85	0.87	0.86
2^{-16}	0.916	0.83	0.86	0.85
2^{-15}	0.925	0.85	0.88	0.86
2^{-14}	0.924	0.84	0.87	0.86
2^{-13}	0.88	0.775	0.825	0.8
2^{-12}	0.89	0.78	0.824	0.81
2^{-11}	0.90	0.78	0.85	0.819
2^{-10}	0.88	0.778	0.83	0.809
2^{-9}	0.9	0.80	8.85	0.824
2^{-8}	0.902	0.81	8.85	0.825
2^{-7}	0.924	0.823	0.87	0.850
2^{-6}	0.925	0.834	0.88	0.86
2^{-5}	0.925	0.85	0.87	0.861
2^{-4}	0.925	0.84	0.87	0.84
2^{-3}	0.93	0.87	0.9	0.88
2^{-2}	0.924	0.83	0.86	0.85
2^{-1}	0.926	0.86	0.88	0.87
2^{0}	0.9	0.84	0.876	0.86
2^{1}	0.9	0.83	0.876	0.84
2^{2}	0.91	0.86	0.878	0.865

In table 2, Different confusion evaluation parameters (such as Acc-Accuracy, Prec-Precision, R-Recall, f-mes- f-measure) are mentioned.

C. Assessment Of Result And Analysis

Six methods of feature selection were compared to authenticate the efficiency of SFR, including Relief-F [7,23], RFS [12], MRMR [16], Fisher Score [6], SVM-RFE-CBR [17], and UGL [19]. We used the same set of parameters for all methods to ensure the experiments were fair, ranging from 10^{-5} to 10^{5}. In both UGL and SVM-RFE-CBR, we used thresholds ranging from 0.6 to 0.9 for highly correlated feature pairs. After 60 features they have been removed from SVM-RFE-CBR, half of them is removed in each iteration until all 60 features have been removed. To run SFR on each set of data, we chose l with set of numbers ranging from 1 to 10 and λ from 0.1 to 1. In our experiments, the number of clustering reps was set to 20.

We used seven different supervised feature selection approaches to pick out various features from each data set in Tables 1. We then performed a 4-fold SVM on the feature set data. Table 3 shows the maximum accuracy against feature selection as per seven approaches on 5 datasets. The proposed method SFR is performed well compared with all other methods in terms of accuracy. We improved 8% of accuracy as SFR on the 14T dataset compared to Relief-F, the runner-up. SFR had the best result on the ST dataset on only twenty features. For the most part, SFR performed admirably across all datasets.

TABLE III. THE ACCURACIES VS THE NUMBER OF SELECTED FEATURES BY DIFFERENT APPROACHES ON 5 DATA SETS.

	Number of selected features→					
Accuracy↓	20	60	100	140	180	220
	Output on the BR3					
SFR	0.68	0.75	0.78	0.76	0.75	0.79
ReF	0.54	0.64	0.65	0.6	0.61	0.6
RFS	0.75	0.73	0.74	0.65	0.62	0.6
MRMR	0.54	0.53	0.54	0.56	0.55	0.52
Fir	0.73	0.69	0.67	0.68	0.72	0.7
SRB	0.6	0.63	0.62	0.62	0.63	0.62
UGL	0.76	0.77	0.72	0.73	0.72	0.62
	Output on the ST data set					
SFR	1	0.97	0.96	0.98	0.99	0.95
ReF	0.97	0.96	0.99	0.99	0.97	0.96
RFS	0.95	0.94	0.92	0.94	0.93	0.95
MRMR	0.73	0.74	0.7	0.68	0.7	0.72
Fir	0.97	0.97	0.96	0.94	0.95	0.96
SRB	0.7	0.92	0.91	0.92	0.95	0.94
UGL	0.95	0.9	0.9	0.91	0.89	0.93
	Output on the BT2 data set.					
SFR	0.85	0.85	0.85	0.86	0.8	0.81
ReF	0.75	0.77	0.81	0.8	0.77	0.8
RFS	0.9	0.72	0.75	0.77	0.77	0.76
MRMR	0.3	0.32	0.4	0.39	0.38	0.35
Fir	0.8	0.84	0.82	0.83	0.84	0.82
SRB	0.63	0.72	0.7	0.68	0.68	0.6
UGL	0.62	0.61	0.61	0.6	0.5	0.51
	Output on the 11T data set					
SFR	0.69	0.87	0.86	0.86	0.89	0.87
ReF	0.75	0.77	0.76	0.76	0.77	0.76
RFS	0.71	0.7	0.7	0.69	0.69	0.69
MRMR	0.3	0.3	0.25	0.3	0.31	0.3
Fir	0.6	0.82	0.9	0.87	0.86	0.87
SRB	0.7	0.72	0.73	0.74	0.73	0.72
UGL	0.7	0.74	0.73	0.72	0.68	0.69
	Output on the 14T data set					
SFR	0.5	0.6	0.62	0.65	0.64	0.65
ReF	0.42	0.47	0.5.5	0.56	0.56	0.53
RFS	0.4	0.41	0.42	0.44	0.48	0.44
MRMR	0.15	0.2	0.19	0.17	0.2	0.19
Fir	0.38	0.5	0.49	0.5	0.51	0.5
SRB	0.4	0.48	0.52	0.55	0.56	0.57
UGL	0.24	0.26	0.38	0.37	0.36	0.35

Three parameters l, η, λ are tested in this experiment to see how they affect SFR's performance. We begin by looking at how l affects SFR's performance. Based on 5 datasets, this shows that the overview of SFC for feature selection does certainly support to choose better features for classification on all datasets, as only one feature cluster yields the lowest accuracy. On most datasets, the accuracy increased with an increase in l, as shown in table 3. The accuracy improved with increasing on the ST and BT2 datasets, which both have incredibly high dimensions. From this table it's easy to see that the classification accuracies were stable at 0.90 when set to 0. We also see that for all datasets, η= 1 yields the lowest accuracy. From Eq. (8), we know that the conventional ranking method degenerates to the structured weighting feature ranking with η = 1. These results demonstrate that using a feature ranking system to classify features improved the process. In real applications, we can use domain knowledge to set the three parameters or use grid search to select the best combination of parameters for a better result.

V. CONCLUSIONS

This paper introduced an SFR approach for cluster ranking features in large volume data. These features were clustered using a SFC approach, and the features within each feature cluster were then ranked independently using SFC's subspace weightings. We proposed a structured weighting feature ranking method to find out feature's rank from various feature clusters as per the SFC subspace weights. Experiments on a collection of various datasets proved the efficiency through algorithm. The SFR was put to the test on a set of large datasets against six other feature ranking methods. According to the proposed model on most datasets, SFR performed better than the other six feature ranking methods. The new method selects features that are both informative and diverse, as demonstrated experimentally. SFC will be improved in the future by introducing new techniques such as ensemble learning. Using SFR in real-world applications is also something we'll be working on in the future.

REFERENCES

[1] Rong Wang, Jintang Bian, Feiping Nie, and Xuelong Li, Unsupervised Discriminative Projection for Feature Selection, IEEE Transactions on Knowledge and Data Engineering, VOL. 34, NO. 2, pp. 942 – 953, February 2022.

[2] G. Chandrashekar and F. Sahin, ''A survey on feature selection methods,'' Comput. Elect. Eng., vol. 40, no. 1, pp. 16–28, Jan. 2014.

[3] Bhuyan H. K., Ravi V. K., Analysis of Sub-feature for Classification in Data Mining, IEEE Transaction on Engineering Management, 2021 (Early Published).

[4] [4] Bhuyan H. K., Chakraborty C, Pani S. K., Ravi V. K., Feature and Sub-Feature Selection for Classification using Correlation Coefficient and Fuzzy model, IEEE Transaction on Engineering Management, May, 2021 (Early Published).

[5] Bhuyan H. K., Kamila N. K., Pani S. K., Individual privacy in data mining using fuzzy optimization, Engineering Optimization, July, 2021 (Early Published).

[6] R. O. Duda, P. E. Hart, and D. G. Stork, Pattern Classification. 2nd Edition, pages 688, Hoboken, NJ, USA: Wiley, 2010.

[7] Yuan Wan , Shengzi Sun , and Cheng Zeng, Adaptive Similarity Embedding for Unsupervised Multi-View Feature Selection, IEEE Transactions On Knowledge And Data Engineering, Vol. 33, No. 10, pp. 3338 – 3350, October 2021.

[8] Jiajing Zhu , Yongguo Liu , Chuanbiao Wen, and Xindong Wu, DGDFS: Dependence Guided Discriminative Feature Selection for Predicting Adverse Drug-Drug Interaction, IEEE Transactions On Knowledge And Data Engineering, Vol. 34, No. 1, pp. 271 – 285, January 2022.

[9] X. Chen, F. Nie, G. Yuan, and J. Z. Huang, ''Semi-supervised feature selection via rescaled linear regression,'' in Proc. 26th Int. Joint Conf. Artif. Intell., pp. 1525–1531, 2017.

[10] Bhuyan H.K., L Raghu Kumar, Reddy K. Ramakrishna: Optimization model for Sub-feature selection in data mining, 2nd International Conference on Smart Systems and Inventive Technology (ICSSIT 2019), IEEE Explore, pp. 1212 - 1216 2019.

[11] M. Masaeli, G. Fung, and J. G. Dy, ''From transformation-based dimensionality reduction to feature selection,'' in Proc. Int. Conf. Mach. Learn., pp. 1–8, 2010,

[12] F. Nie, H. Huang, X. Cai, and C. H. Ding, ''Efficient and robust feature selection via joint `2,1-norms minimization,'' in Proc. Adv. Neural Inf. Process. Syst., pp. 1813–1821, 2010.

[13] D. Wang, F. Nie, and H. Huang, ''Feature selection via global redundancy minimization,'' IEEE Trans. Knowl. Data Eng., vol. 27, no. 10, pp. 2743–2755, Oct. 2015.

[14] X. Chen, J. Z. Huang, Q. Wu, and M. Yang, ''Subspace weighting co-clustering of gene expression data,'' IEEE/ACM Trans. Comput. Biol. Bioinf., vol. 1, no. 99, pp. 1–12, May 2017.

[15] Bhuyan, H. K. and Reddy C. V. Madhusudan: Sub-feature selection for novel classification, International Conference on Inventive Communication and Computational Technologies (ICICCT), IEEE Explore, pp. 477 – 482, 20-21 April, 2018. DOI: 10.1109/ICICCT.2018.8473206

[16] H. Peng, F. Long, and C. Ding, ''Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy,'' IEEE Trans. Pattern Anal. Mach. Intell., vol. 27, no. 8, pp. 1226–1238, Aug. 2005.

[17] K. Yan and D. Zhang, ''Feature selection and analysis on correlated gas sensor data with recursive feature elimination,'' Sens. Actuators B, Chem., vol. 212, pp. 353–363, Jun. 2015.

[18] A. Das, A. Dasgupta, and R. Kumar, ''Selecting diverse features via spectral regularization,'' in Proc. Adv. Neural Inf. Process. Syst., pp. 1583–1591, 2012.

[19] D. Kong, J. Liu, B. Liu, and X. Bao, ''Uncorrelated group lasso,'' in Proc. 30th AAAI Conf. Artif. Intell., pp. 1765–1771, 2016.

[20] G. Govaert and M. Nadif, Co-Clustering: Models, Algorithms and Applications. 256 Pages, Hoboken, NJ, USA: Wiley, 2013. ISBN: 978-1-848-21473-6.

[21] A. Banerjee, I. Dhillon, J. Ghosh, S. Merugu, and D. S. Modha, ''A generalized maximum entropy approach to bregman co-clustering and matrix approximation,'' J. Mach. Learn. Res., vol. 8, pp. 1919–1986, Aug. 2007.

[22] X. Chen, M. Yang, J. Z. Huang, and Z. Ming, ''TWCC: Automated two-way subspace weighting partitional co-clustering,'' Pattern Recognit., vol. 76, pp. 404–415, Apr. 2018.

[23] H. Liu and H. Motoda, Computational Methods of Feature Selection (Chapman & Hall/CRC Data Mining and Knowledge Discovery Series). 1st Edition, 440 pages, London, U.K.: Chapman & Hall, 2007.

[24] Renjie Chen, Ning Sun, Xiaojun Chen, Min Yang, Qingyao Wu. "Supervised Feature Selection with a Stratified Feature Weighting Method", IEEE Access, Volume 6, pp. 15087 – 15098, 2018