

An Unsupervised Feature Selection Algorithm with Feature Ranking for Maximizing Performance of the Classifiers

Danasingh Asir Antony Gnana Singh¹ Subramanian Appavu Alias Balamurugan² Epiphany Jebamalar Leavline³

¹Department of Computer Science and Engineering, Bharathidasan Institute of Technology, Anna University, Tiruchirappalli 620024, India

²Department of Information Technology, K.L.N College of Information Technology, Sivagangai 630611, India

³Department of Electronics and Communication Engineering, Bharathidasan Institute of Technology, Anna University, Tiruchirappalli 620024, India

Abstract: Prediction plays a vital role in decision making. Correct prediction leads to right decision making to save the life, energy, efforts, money and time. The right decision prevents physical and material losses and it is practiced in all the fields including medical, finance, environmental studies, engineering and emerging technologies. Prediction is carried out by a model called classifier. The predictive accuracy of the classifier highly depends on the training datasets utilized for training the classifier. The irrelevant and redundant features of the training dataset reduce the accuracy of the classifier. Hence, the irrelevant and redundant features must be removed from the training dataset through the process known as feature selection. This paper proposes a feature selection algorithm namely unsupervised learning with ranking based feature selection (FSULR). It removes redundant features by clustering and eliminates irrelevant features by statistical measures to select the most significant features from the training dataset. The performance of this proposed algorithm is compared with the other seven feature selection algorithms by well known classifiers namely naive Bayes (NB), instance based (IB1) and tree based J48. Experimental results show that the proposed algorithm yields better prediction accuracy for classifiers.

Keywords: Feature selection algorithm, classification, clustering, prediction, feature ranking, predictive model.

1 Introduction

Prediction plays a vital role in decision making. This is done by the predictive model that is known as classifier or supervised learner^[1]. This predictive model is employed to predict the unseen and unknown data in various applications^[2]. Improving the predictive accuracy is a major challenge among the researchers. The accuracy of the classifier highly relies on the training dataset which is used to train the predictive model. The irrelevant and redundant features of the training dataset reduce the predictive accuracy^[3]. Hence, irrelevant and redundant features must be eliminated from the training dataset for improving the accuracy of the predictive model. The performance improvement of the predictive model include improving the predictive accuracy, reducing the time taken to build the predictive model and reducing the number of features present in the training dataset^[3]. Selecting significant features from the training dataset for improving the performance of the supervised or unsupervised learner is known as feature selection. This feature selection is classified as wrapper, filter, embedded and hybrid methods^[4].

The filter method selects the important features and removes the redundant and irrelevant features from the given

dataset using mathematical statistical measure. It performs well regardless of the classifier chosen, therefore it can work for any type of classification algorithm^[5, 6].

The embedded method uses the training phase of the supervised learning algorithm for selecting the important features from the training dataset. Therefore, its performance is based on the classifiers used^[7].

The wrapper method uses the classifier to determine important features available in the dataset^[8]. The combination of filter and wrapper method is known as hybrid method^[4, 9].

Many feature selection algorithms are proposed recently, which concentrate only on removing the irrelevant features. They do not deal with removing the redundant features from the training dataset. These redundant features reduce the accuracy of the classifiers. This paper proposes a filter based feature selection algorithm named as unsupervised learning with ranking based feature selection (FSULR). This algorithm selects the most significant features from the dataset and removes the redundant and irrelevant features. The unsupervised learner learns the training dataset using expectation maximization function and groups the features into various clusters and these clustered features are ranked based on the χ^2 measure within the cluster. The significant features from each cluster are chosen based on the threshold function.

Regular Paper
Manuscript received November 4, 2013; accepted April 3, 2014
Recommended by Editor-in-Chief Huo-Sheng Hu
© Institute of Automation, Chinese Academy of Science and Springer-Verlag Berlin Heidelberg 2015

The remaining paper is organized as follows. The relevant literature is reviewed in Section 2. Section 3 discusses the proposed feature selection algorithm. Section 4 explores and discusses the experimental results. Conclusion is drawn in Section 5 with future enhancement of this work.

2 Related work

This section discusses various types of feature selection, cluster and classification algorithms as a part of the related works of this proposed algorithm.

2.1 Feature selection algorithm

The feature selection algorithm selects the most significant features from the dataset using ranking based techniques, subset based and unsupervised based techniques. In the ranking based technique, the individual features " f_i " are ranked by applying one of the mathematical measures such as information gain, gain ratio, on the training dataset " TD ". The ranked features are selected as significant features for learning algorithm by a threshold value " T_V " calculated by the threshold function. In subset based technique, the features of the training datasets are separated into maximum number of possible feature subsets " S " and each subset is evaluated by an evaluation criteria to identify the significance of the features present in the subset. The subset containing most significant features is considered as a selected candidate feature subset. In unsupervised based technique, the cluster analysis is carried out to identify the significant features from the training dataset^[10–12].

2.1.1 Feature selection based on correlation (FS-Cor)

In this feature subset selection, the entire feature set $F = \{f_1, f_2, \dots, f_x\}$ of a training dataset " TD " is sub divided into feature subsets " FS_i ". Then, two types of the correlation measures are calculated on each feature subset " FS_i ". One is the feature-feature correlation that is the correlation measure among the features present in a feature subset " FS_i ", and the other one is feature-class correlation that is the correlation measure between the individual feature and the class value of the training dataset. These two correlation measures are computed for all the individual feature subsets of the training dataset. The significant feature subset is identified based on the comparison between the feature-feature correlation and feature-class correlation. If the feature-class correlation value is higher than the feature-feature correlation value, the corresponding feature subset is selected as a significant feature subset of the training dataset^[13].

2.1.2 Feature selection based on consistency measure (FSCon)

In this feature subset selection, entire feature set " F " of the training dataset " TD " is subdivided into maximum possible combinations of feature subsets " FS_i " and the consistency measure is calculated for each subset to identify the significant feature subset. The consistency measure ensures that the same tuple " T_i " is not present in a different class

" C_i "^[5, 14, 15].

2.1.3 Feature selection based on χ^2 (FSChi)

This is a ranking based feature selection technique. The χ^2 statistical measure is applied on the training dataset " TD " to identify the significance level of each feature presents in the training dataset. The χ^2 value " CV " is computed based on the sum of ratio of the difference between the observed ($o_{i,j}$) and expected (e_{ij}) frequencies of the features f_i, f_j to the expected frequencies (e_{ij}) of the features f_i, f_j of the possible instance value combinations of the features^[16].

2.1.4 Feature selection based on information gain (FSInfo)

In this feature selection technique, the information gain measure is applied on the training dataset to identify the significant features based on information gain value of the individual features^[10] in terms of entropy. The entropy value of each feature of the training dataset " TD " is calculated and ranked based on the information gain value^[17].

2.1.5 Feature selection based on gain ratio (FSGai-ra)

In this feature selection technique, the information gain ratio $GR(f)$ is calculated for each feature of the training dataset " TD " to identify the significant feature based on the information present in the features of the " TD "^[17].

2.1.6 ReliefF

This feature selection technique selects the significant features from the training dataset " TD " based on the weighted probabilistic function $w(f)$ with nearest neighbor principle. If the nearest neighbors of a tuple " T " belong to the same class, it is termed as "nearest hit". If the nearest neighbors of a tuple " T " belong to a different class, it is termed as "nearest miss". The probabilistic weight function value $w(f)$ is calculated based on "nearest hit" and "nearest miss" values^[18–20].

2.1.7 Feature selection based on symmetric uncertainty (FSUnc)

This technique uses the correlation measure to select the significant feature from the training dataset " TD ". In addition to that, the symmetric uncertainty " SU " is calculated using the entropy measure to identify the similarity between the two features f_i and f_j ^[21, 22].

2.1.8 Feature selection based on unsupervised learning algorithm

Unsupervised learning is formally known as clustering algorithm. This algorithm groups similar objects with respect to the given criteria like density, distance, etc. The objects present in a group are highly similar than the outliers. This technique is applied for selecting the significant features from the training dataset. Each unsupervised algorithm has its own advantages and disadvantages that determine the application of each algorithm. This paper utilizes expectation maximization (EM) clustering technique to select the significant features by identifying the independent features in order to remove the redundant features from a training dataset " TD "^[23–25].

2.2 Supervised learning algorithm

The supervised learning algorithm builds the predictive model “PM” by learning the training dataset “TD” to predict the unlabeled tuple “T”. This model can be built by various supervised learning techniques such as tree based, probabilistic and rule based. This paper uses the supervised learners namely naive bayes (NB), instance based IB1 and C4.5/J48 supervised learners to evaluate and compare the performance of the proposed feature selection algorithm in terms of predictive accuracy and time taken to build the prediction model with the existing algorithms^[26–28].

2.2.1 Naive Bayes (NB) classifier

This classifier works based on the Bayesian theory. The probabilistic function is applied on training dataset “TD” that contains the features $F = \{f_1, f_2, \dots, f_x\}$, tuple $T = \{t_1, t_2, \dots, t_y\}$ and classes $C = \{c_1, c_2, \dots, c_z\}$. The unlabeled tuple “UT” is predicted to identify its class C_i with the probabilistic condition $P(C_k|T) > P(C_w|T)$ where $k \neq w$ ^[29].

2.2.2 Decision tree based C4.5/J48 classifier

This tree based classifier constructs the predictive model using decision tree. Basically, the statistical tool information gain is used to learn the dataset and construct the decision tree. Information gain is computed for each attribute present in the training dataset “TD”. The feature with higher information value is considered as the root node and the dataset is divided into further levels. The information gain value is computed for all the nodes and this process is iterated until a single class value is obtained in all the nodes^[30, 31].

2.2.3 IB1

This classifier utilizes the nearest neighbor principle to measure the similarity among the objects to predict the unlabeled tuple “UT”. The Euclidean distance measure is used to compute the distance between various tuples present in the training dataset^[32].

3 Unsupervised learning with ranking based feature selection (FSULR)

This algorithm follows a sequence of steps to select the significant features from the training dataset. Initially, the training dataset is transposed. Then features are clustered and the features in each cluster are ranked based on the χ^2 value. Then the threshold value is computed to select highly significant features from each clustered features. All the selected features from different clusters are combined together as candidate significant features selected from the training dataset.

Algorithm 1. FSULR

Input: Training dataset TD, tuple set $T = \{t_1, t_2, \dots, t_l\}$, feather set $F = \{f_1, f_2, \dots, f_m\}$, class labels $C = \{C_1, C_2, \dots, C_n\}$ and $T, F, C \in TD$

Output: Selected significant feature subset $SSF = \{f_{1,2}, \dots, f_i\} \in F$

Method:

- 1) Begin
- 2) Swap (TD)
- 3) {Return → Swapped $TD = SD$ } // Swap the tuple-T (row) into feature-F (column)
- 4) EM_Cluster (SD)
- 5) {Return # of grouped feature sets $GF = \{GF_1, GF_2, GF_3, \dots, GF_n\}$ } // $GF \in F$
- 6) for ($L = 1, L \leq n, L++$)
- 7) $\{\chi^2(GF_L, C)$
- 8) {Return ranked feature members of grouped feature set RGF_K , where $K = 1, \dots, n$ with χ^2 value CV} // Computing and ranking the χ^2 value for the grouped features
- 9) Threshold(CV_Max, CV_Min) // CV_Max, CV_Min are maximum and minimum χ^2 values, respectively
- 10) {Return the value “ φ ”} // φ is break-off threshold value
- 11) for ($M = 1, M \leq n, M++$)
- 12) { Break_off_Feature (RGF_M, CV_M, φ)
- 13) { Return BFSS_K, where $K = 1, \dots, n$ } // BFSS_K is the break off feature subset based on CV
- 14) for ($N = 1, N \leq n, N++$)
- 15) {Return SSF=Union(BFSS_N)} // SSF is the selected significant feature subset
- 16) End.

Phase 1. In this phase, the FSULR algorithm receives the training dataset “TD” as input with the feature $F = \{f_1, f_2, \dots, f_x\}$, tuples $T = \{t_1, t_2, \dots, t_y\}$ and class $C = \{c_1, c_2, \dots, c_z\}$. Then the class C is removed. The function Swap(·) swaps the features “F” into tuples “T” and returns swapped dataset “SD” for grouping the features “ f_i ”.

Phase 2. In this phase, expectation maximization maximal likelihood function^[33] and Gaussian mixture model as seen in (1) and (2) are used to group the features using the function EM-Cluster(·). It receives the swapped dataset “SD” to group the “SD” into grouped feature sets $GF = \{GF_1, GF_2, GF_3, \dots, GF_n\}$.

$$\eta(b|Dz) = P(Dz|b) = \prod_{k=1}^m P(z_k|b) = \prod_{k=1}^m \sum_{l \in L} P(b_k|L, b)P(L|b) \tag{1}$$

where D denotes the dataset containing an unknown parameter b and known product Z and L is the unseen data. Expectation maximization is used as distribution function and this is derived as Gaussian mixture model multivariate distribution $P(z|L, b)$ as expressed in (2).

$$P(z|L, b) = \frac{1}{(2\pi)^{\frac{\alpha}{2}} \sqrt{\Delta(\sum_l)}} e^{-\frac{1}{2}(z-\lambda_l)^T \Sigma_l^{-1} (z-\lambda_l)} \tag{2}$$

With this scenario, a value α is randomly picked from the set “S” with instances “I” from non-continuous set $N = \{1, 2, 3, \dots, n\}$. Let λ_l be the mean vector of the unseen variable and the covariance matrix $\sum_l (l = 1, 2, \dots, n)$

with the discrete distribution $P(l|b)$. The cross-validation computes the likelihood in correspondence to the value n to determine n Gaussian components for computing the expectation maximization. This developed Gaussian mixture model^[34] is used for clustering the features based on the level of Gaussian component possessed by the individual features " f_i "^[35] and return the n number of grouped feature subsets $GF = \{GF_1, GF_2, GF_3, \dots, GF_n\}$.

Phase 3. In this phase, the most significant features are selected from the clustered feature subset by $\chi^2(\cdot)$ function. It receives the clustered feature set GF_K with the class feature C and computes the χ^2 value " CV " for all the features as shown in (3)^[16]

$$CV = \sum_{i=1}^m \sum_{j=1}^n \frac{(o_{ij} - e_{ij})}{e_{ij}}. \quad (3)$$

The feature " f_i " contains distinct tuple values tv_1, tv_2, \dots, tv_n . Then, the χ^2 value " CV " is computed for the features f_i and f_j . o_{ij} is the observed frequency and e_{ij} is the expected frequency of possible tuple values. To identify the significant features, the features are ranked by their χ^2 value " CV " and the ranked grouped feature set is obtained. The Threshold(\cdot) function receives the CV_Min and CV_Max values and returns the break-off threshold value φ as shown in (4). The function Break_off_Feature(\cdot) recognizes the ranked grouped feature set RGF with φ and breaks off the top ranked " CV " up to the value φ and returns them as break off feature subset BFSS.

$$\varphi = H \times \left(\frac{\alpha - \beta}{2} + \beta \right) \quad (4)$$

where φ is break off-threshold value, H is threshold constant that is equal to 1, α is the maximum χ^2 value (CV_Max) of the feature, and β is the minimum χ^2 value (CV_Min) of the feature.

Phase 4. This phase combines break off feature subset BFSS threshold from the clustered features by the function Union(\cdot) and produces the selected significant feature subset SSF as the candidate selected features.

4 Experimental setup

Totally, 21 datasets are used to conduct the experiment with number of features ranging from 2 to 684, number of instances from 5 to 1728 and number of class labels from 2 to 24 as listed in the Table 1. These datasets are taken from the Weka software^[36] and UCI Repository^[37]. The performance of the proposed algorithm FSULR is analyzed and compared with the other feature selection algorithms FSCor, FSChi, FSCon, FSGai-ra, ReliefF, FSUnc and FSInfo with NB, J48 and IB1 classifiers.

5 Experimental procedure, results and discussion

This experiment is conducted with 21 well known publicly available datasets as shown in the Table 1 by 7 feature selection algorithms namely FSCor, FSChi, FSCon, FSGai-ra, ReliefF, FSUnc and FSInfo. Initially, the datasets are fed as an input to the feature selection algorithms and the selected features are obtained as output of the feature selection algorithms. These selected features are given to the

Table 1 Datasets

Serial number	Dataset	Number of features	Number of instances	Number of classes
1	Breast cancer	9	286	2
2	Contact lenses	24	5	5
3	Credit-g	20	1000	2
4	Diabetes	8	768	2
5	Glass	9	214	7
6	Ionosphere	34	351	2
7	Iris 2D	2	150	3
8	Iris	4	150	3
9	Labor	16	57	2
10	Segment challenge	19	1500	7
11	Soybean	683	36	14
12	Vote	435	17	3
13	Weather nominal	4	14	2
14	Weather numeric	14	5	2
15	Audiology	69	226	24
16	Car	6	1728	4
17	Cylinder bands	39	540	2
18	Dermatology	34	366	3
19	Ecoli	7	336	8
20	Anneal	39	898	6
21	Hay-train	10	373	9

NB, J48 and IB1 classifiers and the predictive accuracy and the time taken to build the predictive model are calculated with the 10-fold cross validation test mode.

The performance of the proposed FSULR algorithm is analyzed in terms of predictive accuracy, time to build the model and the number of features reduced. Fig. 1 expresses the overall performance of feature selection algorithms in producing the predictive accuracy for supervised learners and it is observed that the FSULR performs better than all the other feature selection algorithms compared. The second and third positions are retained by the FSCon and FSInfo respectively. Fig. 2 shows that the proposed FSULR achieves better accuracy than the other algorithms compared for NB and IB1 classifiers. The FSCon achieves better results for J48 classifier as compared to all the other classifiers.

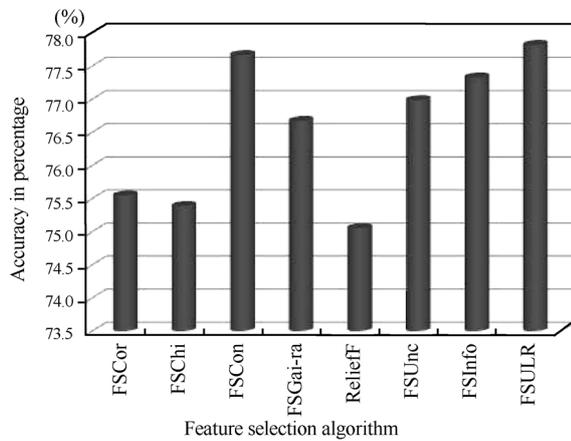


Fig. 1 Comparison of overall prediction accuracy in percentage with respective feature selection algorithm

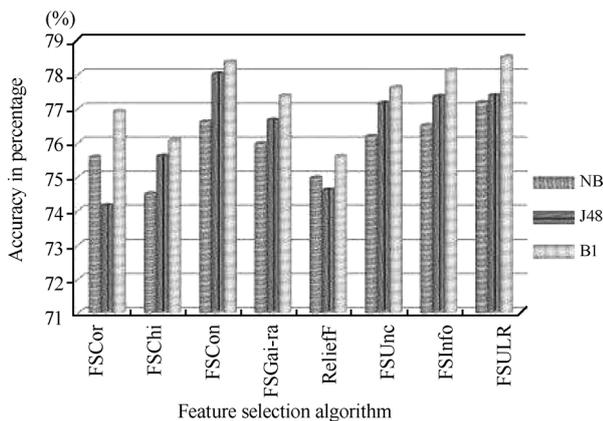


Fig. 2 Comparison of prediction accuracy in percentage with respect to the feature selection algorithm and classifier

From Fig.3, it is evident that FSULR takes more time to build the predictive model compared to all other algorithms. From Fig.4, it is observed that the FSGai-ra and FSUnc require less time to build the model for NB classifier than other algorithms compared. The FSUnc takes less time to build model for J48 classifier. The FSGai-ra and

ReliefF consume less time to build the predictive model for IB1 classifier. Fig.5 exhibits that ReliefF reduces the number of features significantly than other algorithms compared and FSCon reduces the least number of features.

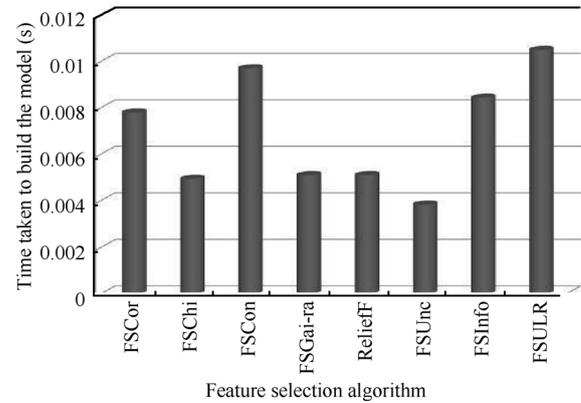


Fig. 3 Comparison of overall time taken to build the predictive model (in second) with respect to the feature selection algorithm

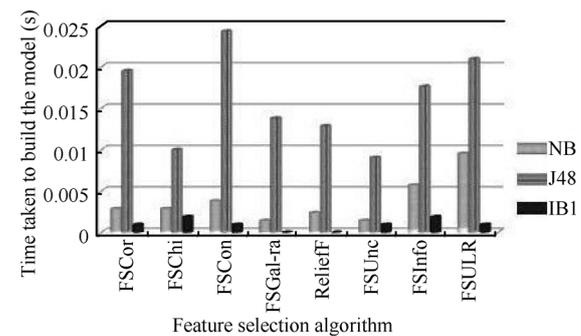


Fig. 4 Comparison of time taken to build the predictive model (in second) with respect to the feature selection algorithm and classifier

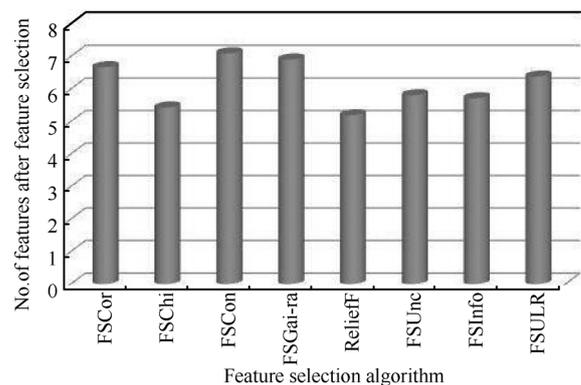


Fig. 5 Comparison of number of features reduction with respect to the feature selection algorithm

6 Conclusions and future work

This paper proposed a feature selection algorithm namely unsupervised learning with ranking based feature selection algorithm (FSULR). Performance of this algorithm is an-

alyzed in terms of accuracy produced for classifier, time taken to build predictive model and feature reduction. The performance of this algorithm is compared with other feature selection algorithms namely correlation based FSCor, χ^2 based FSChi, consistency based FSCon, gain ratio based FSGai-ra, ReliefF, symmetric uncertainty based FSUnc and information gain based FSInfo algorithm with NB, J48, IB1 classifiers. The FSULR achieves better prediction accuracy than the other feature selection algorithms and achieves higher accuracy for IB1 and NB classifiers compared to the other feature selection algorithms. The FSULR is also considerably good in reducing the time to build model for IB1 compared to FSGai-ra and ReliefF. FSULR is considerably good in reducing the time to build model for J48 classifiers compared to FSCon algorithm. The FSULR reduces the number of features compared to FSCor, FSCon and FSGai-ra. In future, this work can be extended with other statistical measures for ranking the features within the clusters.

References

- [1] J. Sinno, Q. Y. Pan. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1355, 2010.
- [2] M. R. Rashedur, R. M. Fazole. Using and comparing different decision tree classification techniques for mining ICDDR, B Hospital Surveillance data. *Expert Systems with Applications*, vol. 38, no. 9, pp. 11421–11436, 2011.
- [3] M. Wasikowski, X. W. Chen. Combating the small sample class imbalance problem using feature selection. *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1388–1400, 2010.
- [4] Q. B. Song, J. J. Ni, G. T. Wang. A fast clustering-based feature subset selection algorithm for high-dimensional data. *IEEE Transactions on Knowledge and Data Engineering*, vol. 15, no. 1, pp. 1–14, 2013.
- [5] J. F. Artur, A. T. Mário. Efficient feature selection filters for high-dimensional data. *Pattern Recognition Letters*, vol. 33, no. 13, pp. 1794–1804, 2012.
- [6] J. Wu, L. Chen, Y. P. Feng, Z. B. Zheng, M. C. Zhou, Z. H. Wu. Predicting quality of service for selection by neighborhood-based collaborative filtering. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 43, no. 2, pp. 428–439, 2012.
- [7] C. P. Hou, F. P. Nie, X. Li, D. Yi, Y. Wu. Joint embedding learning and sparse regression: A framework for unsupervised feature selection. *IEEE Transactions on Cybernetics*, vol. 44, no. 6, pp. 793–804, 2014.
- [8] P. Bermejo, L. dela Ossa, J. A. Gámez, J. M. Puerta. Fast wrapper feature subset selection in high-dimensional datasets by means of filter re-ranking Original Research Article. *Knowledge-based Systems*, vol. 25, no. 1, pp. 35–44, 2012.
- [9] S. Atulji, G. Shameek, V. K. Jayaramanb. Hybrid biogeography based simultaneous feature selection and MHC class I peptide binding prediction using support vector machines and random forests. *Journal of Immunological Methods*, vol. 387, no. 1-2, pp. 284–292, 2013.
- [10] H. L. Wei, S. Billings. Feature subset selection and ranking for data dimensionality reduction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 1, pp. 162–166, 2007.
- [11] X. V. Nguyen, B. James. Comments on supervised feature selection by clustering using conditional mutual information-based distances. *Pattern Recognition*, vol. 46, no. 4, pp. 1220–1225, 2013.
- [12] A. G. Iffat, S. S. Leslie. Feature subset selection in large dimensionality domains. *Pattern Recognition*, vol. 43, no. 1, pp. 5–13, 2010.
- [13] M. Hall. Correlation-based Feature Selection for Machine Learning, Ph.D dissertation, The University of Waikato, New Zealand, 1999.
- [14] Y. Lei, L. Huan. Efficient feature selection via analysis of relevance and redundancy. *Journal of Machine Learning Research*, vol. 5, no. 1, pp. 1205–1224, 2004.
- [15] M. Dash, H. Liu, H. Motoda. Consistency based feature selection. In *Proceedings of the 4th Pacific Asia Conference on Knowledge Discovery and Data Mining*, Kyoto, Japan, pp. 98–109, 2000.
- [16] H. Peng, L. Fulmi, C. Ding. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, 2005.
- [17] H. Uğuz. A two-stage feature selection method for text categorization by using information gain, principal component analysis and genetic algorithm. *Knowledge-based Systems*, vol. 24, no. 7, pp. 1024–1032, 2011.
- [18] M. Robnik-Šikonja, I. Kononenko. Theoretical and empirical analysis of ReliefF and RReliefF. *Machine Learning*, vol. 53, no. 1-2, pp. 23–69, 2003.
- [19] S. Yijun, T. Sinisa, G. Steve. Local-learning-based feature selection for high-dimensional data analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1610–1626, 2010.
- [20] W. Peng, S. Cesar, S. Edward. Prediction based on integration of decisional DNA and a feature selection algorithm RELIEF-F. *Cybernetics and Systems*, vol. 44, no. 3, pp. 173–183, 2013.

- [21] H. W. Liu, J. G. Sun, L. Liu, H. J. Zhang. Feature selection with dynamic mutual information. *Pattern Recognition*, vol. 42, no. 7, pp. 1330–1339, 2009.
- [22] M. C. Lee. Using support vector machine with a hybrid feature selection method to the stock trend prediction. *Expert Systems with Applications*, vol. 36, no. 8, pp. 10896–10904, 2009.
- [23] P. Mitra, C. A. Murthy, S. K. Pal. Unsupervised feature selection using feature similarity. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 3, pp. 301–312, 2002.
- [24] J. Handl, J. Knowles. Feature subset selection in unsupervised learning via multi objective optimization. *International Journal of Computational Intelligence Research*, vol. 2, no. 3, pp. 217–238, 2006.
- [25] H. Liu, L. Yu. Toward integrating feature selection algorithms for classification and clustering. *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 4, pp. 491–502, 2005.
- [26] S. García, J. Luengo, J. A. Sáez, V. Lóez, F. Herrera. A survey of discretization techniques: Taxonomy and empirical analysis in supervised learning. *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 4, pp. 734–750, 2013.
- [27] S. A. A. Balamurugan, R. Rajaram. Effective and efficient feature selection for large-scale data using Bayes' theorem. *International Journal of Automation and Computing*, vol. 6, no. 1, pp. 62–71, 2009.
- [28] J. A. Mangai, V. S. Kumar, S. A. alias Balamurugan. A novel feature selection framework for automatic web page classification. *International Journal of Automation and Computing*, vol. 9, no. 4, pp. 442–448, 2012.
- [29] H. J. Huang, C. N. Hsu. Bayesian classification for data from the same unknown class. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetic*, vol. 32, no. 2, pp. 137–145, 2002.
- [30] S. Ruggieri. Efficient C4.5. *IEEE Transactions on Knowledge and Data Engineering*, vol. 14, no. 2, pp. 438–444, 2002.
- [31] P. Kemal, G. Salih. A novel hybrid intelligent method based on C4.5 decision tree classifier and one-against-all approach for multi-class classification problems. *Expert Systems with Applications*, vol. 36, no. 2, pp. 1587–1592, 2009.
- [32] W. W. Cheng, E. Hüllermeier. Combining instance-based learning and logistic regression for multilabel classification. *Machine Learning*, vol. 76, no. 3, pp. 211–225, 2009.
- [33] J. H. Hsiu, P. Saumyadipta, I. L. Tsung. Maximum likelihood inference for mixtures of skew Student-*t*-normal distributions through practical EM-type algorithms. *Statistics and Computing*, vol. 22, no. 1, pp. 287–299, 2012.
- [34] M. H. C. Law, M. A. T. Jain, A. K. Jain. Simultaneous feature selection and clustering using mixture models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 9, pp. 1154–1166, 2004.
- [35] T. W. Lee, M. S. Lewicki, T. J. Sejnowski. ICA mixture models for unsupervised classification of non-gaussian classes and automatic context switching in blind signal separation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1078–1089, 2002.
- [36] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann. The WEKA data mining software: An update. *ACM SIGKDD Explorations Newsletter*, vol. 11, no. 1, pp. 10–18, 2009.
- [37] K. Bache, M. Lichman. *UCI Machine Learning Repository*, [Online], Available: <http://archive.ics.uci.edu/ml>, Irvine, CA: University of California, School of Information and Computer Science, 2013.



Danasingh Asir Antony Gnana Singh received the M.Eng. and B.Eng. degrees from Anna University, India. He is currently working as a teaching fellow in the Department of Computer Science and Engineering, Bharathidasan Institute of Technology, Anna University, India.

His research interests include data mining, wireless networks, parallel computing, mobile computing, computer networks, image processing, software engineering, soft computing, teaching learning process and engineering education.

E-mail: asirantony@gmail.com (Corresponding author)
ORCID iD: 0000-0002-4913-2886



Subramanian Appavu Alias Balamurugan received the Ph.D. degree from Anna University, India. He is currently working as a professor and the head of the Department of Information Technology in K.L.N College of Information Technology, India.

His research interests include pattern recognition, data mining and informatics.

E-mail: app_s@yahoo.com



Epiphany Jebamalar Leavline received the M.Eng. and B.Eng. degrees from Anna University, India, and received the MBA degree from Alagappa University, India. She is currently working as an assistant professor in the Department of Electronics and Communication Engineering, Bharathidasan Institute of Technology, Anna University, India.

Her research interests include image processing, signal processing, VLSI design, data mining, teaching learning process and engineering education.

E-mail: jebi.lee@gmail.com