DOI: http://dx.doi.org/10.26483/ijarcs.v8i7.4222

ISSN No. 0976-5697

Volume 8, No. 7, July – August 2017

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

EFFICIENT FEATURE SELECTION TECHNIQUE BASED ON MODIFIED FUZZY C-MEANS CLUSTERING WITH ROUGH SET THEORY

P.Arumugam
Professor,Department of Statistics
Manonmaniam Sundaranar University
Tirunelveli, 628012, Tamilnadu
India

P.Jose
Research scholar, Department of Statistics
Manonmaniam Sundaranar University
Tirunelveli, 628012, Tamilnadu
India

Abstract;Feature selection plays an important role in classification, since it can shorten the learning time, simplify the learning classifiers, and improve the classification performance. There may be complex interaction among features; it is generally difficult to find the best feature subset. This article presents an efficient feature selection based on Modified Fuzzy c-Means clustering with Rough Set Theory (MFCM-RST), the classification will be done based on the SVM classifier. The proposed algorithm involves the amalgamation of concepts of rough sets, fuzzy sets, and c-Means clustering algorithm. While the fuzzy set enables efficient handling of overlapping partitions, the concept of rough set deals with uncertainty, vagueness, incompleteness, and indiscernibility in class definition. Whereas, the kernel trick projecting the feature space into a higher dimension using an appropriate non-linear mapping function ensures linear separability of the complex clusters which are otherwise not linearly separable in its original feature space. Finally, finds the near optimum values of the different parameters used in the proposed method. The effectiveness of the proposed algorithm is evaluated using a UCI machine learning datasets. Experimental results justify the superiority of the proposed method in comparison to other traditional techniques.

Key words: feature selection Fuzzy c-means, Rough set theory, support vector machine

1. INTRODUCTION

Feature selection plays an important role in classification, since it can shorten the learning time, simplify the learning classifiers, and improve the classification performance. One of the most widely used prototype-based partitional clustering algorithms is Hard c-Means (HCM) [1], where each object must be assigned to exactly one cluster. On the other hand, fuzzy c-means (FCM) relaxes this requirement by allowing gradual memberships [2]. In effect, it offers the opportunity to deal with the data that belong to more than one cluster at the same time. It assigns memberships to an object, which is inversely related to the relative distance of the object to cluster. Although FCM is a very useful clustering method, the resulting membership values do not always correspond well to the degrees of belonging of the data, and it may be inaccurate in a noisy environment [3]. In real-data analysis, noise and outliers are unavoidable. Hence, to reduce this weakness of the FCM and to produce memberships that have a good explanation of the degrees of belonging for the data, Krishnapuram and Keller [4] proposed a possibilistic approach to clustering which used a possibilistic type of membership function in describing the degree of belonging. However, the possibilistic c-means (PCM) sometimes generates coincident clusters [5].

2. RELATED WORK

Recently, the use of both fuzzy (probabilistic) and possibilistic memberships in a clustering algorithm has been proposed in [6]. Rough-set theory [7] is a new paradigm that is used to deal with uncertainty, vagueness, and incompleteness. It has been applied to fuzzy-rule extraction, reasoning with uncertainty, fuzzy modeling, etc. In [8], Lingras and West introduced a new clustering method called rough c-means (RCM), which describes a cluster by a prototype (center) and a pair of lower and upper approximations. The lower and upper approximations are different weighted parameters that are used to compute the new centers. Combination of fuzzy and rough sets provides an important direction in reasoning with uncertainty [9]. Both fuzzy and rough sets provide a mathematical framework to capture uncertainties that are associated with the data . They are complementary in some aspects. Recently, combining both rough and fuzzy sets, Mitra et al. [10] proposed a new c-means algorithm; where each cluster is consist of a fuzzy lower approximation and a fuzzy boundary. Each object in lower approximation takes a distinct weight, which is its fuzzy membership value. However, the objects in lower approximation of a cluster should have a similar influence on the corresponding centroid and cluster, and their weights should be independent of other centroids and clusters. Thus, the concept of fuzzy lower approximation, which is introduced in, reduces the weights of objects of lower approximation. In effect, it drifts the cluster prototypes from their desired locations. Moreover, it is sensitive to outliers. In[11]which reduces the computational complexity, incr eases the clusterUniqueness, and retains the originality of the data using clustering via rough set. In this phase, an optimal subset of features which are necessary and sufficient for solving a problem is selected [12]. Feature selection improves the accuracy of algorithms by reducing the dimensionality and removing irrelevant features [13]. In this paper, breast cancer diagnosis based on a SVM-based method combined with feature selection has been proposed [14] .Conventional Principal Component Analysis (PCA) is one of the most frequently applied feature extraction techniques. It is based on extorting the axes on which data illustrates the maximum changeability [15]. More over Cluster analysis is a commonly used data mining technique to explore the relationships among attributes, samples and the relationships between attributes and samples. Clustering algorithms assign samples or attributes to clusters based on their similarity. Cluster analysis can be used as a preliminary method for classification or for finding new classes. Hierarchical clustering tree (HCT) [16] and k-means [17] are the two most popular clustering methods used to extract the features from the data. However, the rough sets offer an effective approach of managing uncertainties and can be employed for tasks such as data dependency analysis, feature identification, dimensionality reduction, and pattern classification. Rough set theory [18] is a fairly recent intelligent technique for managing uncertainty that is used for the discovery of data dependencies, to evaluate the importance of attributes, to discover patterns in data, and to reduce redundancies. While wrappers and embedded methods require a frequent classifier interaction in their flow, filters do not need any classifier interaction during the construction of the feature set [19]. Effective hybrid attribute reduction algorithm based on a generalized fuzzy rough model. A theoretic framework of fuzzy rough model based on fuzzy relations is accessible, which contain a foundation for algorithm construction. Though, several attributes derive significance measures based on the proposed fuzzy-rough model and create a forward greedy algorithm for hybrid attribute reduction [20]. A feature selection method based on mutual information is proposed in [21] to select a set of genes from microarray gene expression data. It employed a fuzzy rough set theory to calculate the significance and implication of the genes that enormously decreases the computational complication while keeping high predictive accuracy. Moreover, the proposed approach is compared with that of conventional fuzzy rough approach and rough set method using the predictive accuracy of support vector machine and 1-nearest neighbour rule on dissimilar microarray data sets. A structure for fuzzy rough set based feature selection built up around the formal notion of a fuzzy decision reduct was given in [22]. The attribute subset should retain the quality of the feature set to a certain extent, which are able to create shorter attribute subsets. Moreover, they have provided a comprehensive classification evaluation measures that can be used to define fuzzy decision reduct. Jensen [23] summarized the new approaches to fuzzy rough feature selection that are capable of dealing with imprecision and uncertainty. Consequently, it is desirable to hybridize and extend the data imperfection. Such developments offer a high degree of flexibility and provide robust solutions and advanced tools for data analysis. Fuzzy-Rough Set (FRS) based feature selection was exposed to be extremely useful at reducing data dimensionality, but possesses numerous problems that renders it ineffective for large datasets. Moreover, the paper proposed three new methods for fuzzy rough feature selection based on fuzzy similarity relations. Richard Jensen and Qiang Shen [24] summarized the approach based on fuzzy rough set and fuzzy rough feature selection, which can address the retain dataset semantics and problems. Additionally, it can be applied two challenging domains in feature reduction, namely, complex system monitoring and classification. Consequently, this approach was demonstrated and compared with numerous dimensionality reducers. In [25] proposed a wasp swarm optimization algorithm for attribute reduction based on rough set and the consequence of the feature. Moreover, it's based on the mutual information between decision attributes and conditional attributes. Then, the algorithm dynamically calculates heuristic information based on the implication of feature to guide search.

3. PROPOSED METHODOLOGY

In this paper input dataset will be as high dimensional or high features which are a great barrier for classification. Therefore, feature dimension reduction method will be applied to reduce the features' space without losing the accuracy of classification. Here, clustering with rough set theory based feature selection will be developed and used to reduce the feature dimension. In this effort Modified Fuzzy C-Means clustering algorithm with Rough Set Theory (MFCM-RST) will be combined and used to feature selection process of high dimensional data .once the feature reduction is formed Classification done by using SVM.

3.1 Pre-processing

The pre-processing will be applied to extract useful data and to convert suitable sample from raw datasets. In This Paper UCI machine learning repository data set is pre-processed. The dataset contains high dimensional data. Data pre-processing generally the experimental data set usually has some objects which have uncertain value or missing value. Firstly normalize the experimental data sets after eliminating the missing objects before the clustering analysis.

3.2 Feature Selection using Modified Fuzzy c-means algorithm

Among the fuzzy clustering method, the fuzzy c-means (FCM) algorithm is the most well-known method because it has the advantage of robustness for ambiguity and maintains much more information than any hard clustering methods. The algorithm is an extension of the classical and the crisp k-means clustering method in fuzzy set domain. It is widely studied and applied in pattern recognition, image segmentation and image clustering, data mining, wireless sensor network and so on. The basic idea of this section is to reduce the dimension of the features using fuzzy c-means clustering with roughest theory (FCM-RS) algorithm**.** The high number of features is a great obstacle for the prediction; so we have to reduce the dimension of the features. As a result, feature dimension reduction method is required to decrease the features' space without losing the precision of prediction. In addition, we reduce the number of features and remove the not related, unnecessary or noisy information. Besides, this improves the presentation of information prediction with speeding up the processing algorithm. To improve the prediction accuracy, use the FCM-RS algorithm for feature reduction.

3.3 Steps for MFCM –RST Algorithm

Step1: *Subset formation*

The fuzzy c-means (FCM) algorithm is a clustering algorithm developed by Dunn, and later on improved by Bezdek. It is useful when the required numbers of clusters are pre-determined; thus, the algorithm tries to put each of the data points to one of the clusters. What makes FCM different is that it does not decide the absolute membership of a data point to a given cluster; instead, it calculates the likelihood (the degree of membership) that a data point will belong to that cluster. Hence, depending on the accuracy of the clustering that is required in practice, appropriate tolerance measures can be put in place. Since the absolute membership is not calculated, FCM can be extremely fast because the number of iterations required to achieve a specific clustering exercise corresponds to the required accuracy. Clustering techniques are mostly unsupervised methods that can be used to organize data into groups based on similarities among the individual data items. Most clustering algorithms do not rely on assumptions common to conventional statistical methods, such as the underlying statistical distribution of data, and therefore they are useful in situations where little prior knowledge exists. The potential of clustering algorithms to reveal the underlying structures in data can be exploited in a wide variety of applications, including classification, image processing, pattern recognition, modelling and identification. This chapter presents an overview of fuzzy clustering algorithms based on the c means functional. The FCM algorithm assigns cluster centre to each category by using fuzzy memberships.

$$J_m = \sum_{i=1}^{I}\sum_{j=1}^{J}(\mu_{ij})^m \left\| x_i - z_j \right\|^2 \qquad (1)$$

In Eqn. (1), x_i represents the features $t(w(s_n^i))$, $c(s_n^i)$ extracted from the input database, z_j is the j th cluster centre and m is the constant value. The membership function represents the probability that a cluster center belongs to a specific cluster. In the FCM algorithm, the probability is dependent on the distance between the pixel and each individual cluster center in the feature domain. The membership functions and cluster centers are updated by the equations (2) and (4).

$$u_{ij} = \frac{1}{\sum_{k=1}^{J}\left(\frac{\left\|x_i - z_j\right\|}{\left\|x_i - z_k\right\|}\right)^{\frac{2}{m-1}}} \qquad (2)$$

Repeat the algorithm until the coefficients' change between two iterations is no more than ξ, for the given sensitivity threshold.

$$\max_{ij}\left\| U_{ij}^{(k)} - U_{ij}^{(k+1)} \right\| < \xi \qquad (3)$$

In equation (3), ξ is a termination criterion between 0 and 1, whereas k are the iteration steps. The clusters centroid values are computed by using the equation (4).

$$z_j = \frac{\sum_{i=1}^{I} u_{ij}^m . x_i}{\sum_{i=1}^{I} u_{ij}^m} \qquad (4)$$

To enhance the performance of the fuzzy-C-means clustering method, adaptiveness is invoked by measuring the Clustering effectiveness (α) and Absolute density (β). On the basis of these two, we set two thresholds to ensure the clustering being good. After the FCM process, we obtain the number of cluster set such as $I_1, I_2, I_3, ..., I_n$. This clustered dataset is used for the further processing. The function fcm takes a data set and a desired number of clusters and returns optimal cluster centers and membership grades for each data point. You can use this information to build a fuzzy inference system by creating membership functions that represent the fuzzy qualities of each cluster.

- Find 2 clusters using fuzzy c-means clustering.
- Classify each data point into the cluster with the largest membership value.
- Plot the clustered data and cluster center

Tune the 3 Optional Parameters

- Exponent,
- Maximum number of iterations
- Minimum amount of improvement

The function FCM takes a high dimensional data set and a desired number of clusters and returns optimal cluster centers and membership grades for each data point.

Step 2: *Attribute selection*

After the clustering process using FCM, we have to do the attribute selection process. Here, we calculate the minimum $A_{\min}$ and maximum $A_{\max}$ value of each column (or each attribute) in the dataset D_1 and D_2. If the $A_{\min}$ and $A_{\max}$ value is similar to corresponding dataset I_1 and I_2, then have to neglect that column.

Step 3: *Discretization*

Discretization is a significant step in data processing to convert the data into specific interval, means that the range of values is confined into a specific interval. Here, used one discretization function based on the predictable way. Perform the discretization process at first, then identify the maximum and minimum values of every attribute, and the K interval is tracked by taking the ratio between the deviated value and the K value.

$$for\,each\ "j"\ Dev^{j} = \left[\frac{\max(A_j) - \min(A_j)}{3} \right] \qquad (5)$$

Step 4: *Reduct and Core analysis*

Rough set provides a method to decide the importance and necessity of features. It is an extension of set theory proposed by Pawlak for knowledge discovery in data sets [2]. Given a Dataset with discretized features, it is possible to find a reduct of original features that are most predictive in terms of classification accuracy. Clustering is an unsupervised classification that partitions an input data set into a desired number of subgroups or clusters. It is a process of dividing the data points into their natural groups so that the data in the same group or cluster are similar to one another and different from the data points in other groups In Rough set theory Reduct and core are the two most important concept .Reduct is a reduced subset of original set which retains the accuracy of original set. There are usually several such subsets of attributes and those which are minimal are called reducts. Reduct is often used in the attribute selection process to reduce unnecessary attributes towards decision making applications

3.4 Classification using SVM

The input data space mapping to feature space as nonlinear fashion and the mapping caring out from the feature space to estimation phase as linear fashion. Nonlinear kernel for transformation of input training sample data often known as kernel trick. This is a major elegant trick of SVM. To apply linear classification technique to nonlinear classification problems SVM adopts different breed of kernel are polynomial learning machine, radial basis function network, two layer perception, linear kernel. Generally after the process of transformation of kernel.

SVM structured as a two class problem, where the classes are separable linearly. The input dataset D be represent in 2D as (x_1, y1), (x_2, y2)…. ($x_{|D|}$, $y_{|D|}$), where x_i is the set of training tuples and y_i is the class label associated with training sample. In a training sample SVM constructs a line of separation for two attributes (x, y) and a plane of separation for three attributes and a hyper plane of separation for n dimensions. In this environment find and select the best hyper plane that is the maximum marginal hyper plane is the challenging task. Due to this contribution SVM handle in three phases as input phase, learning phase, and function estimation phase. During the learning or training phase SVM searches the largest margin hyper plane between classes. The construction of an optimal hyper plane for linearly separable patterns is not performing well. Which is continuously considered an extended in a principled way to deal with optimal hyper plane for no separable patterns.SVM builds inner product kernel method. It produces space from the training sample into support vector xi drawn from the input data space. The support vector represents a small subset of data points. The data point in the hyper plane is called margin of separation. The data points on outer margin and inner margin are on boundary lines are specified as support vectors. Kernel allows SVM to fit maximum margin hyper plane in the feature space. This is perhaps on transformation phase. The transformation may be non-linear. Most probably the kernel method design is optimal in SVM this optimality being rooted in convex optimization. Multidimensional linear decision surface in the input space is represented in the form of

$$w^{T}_{o}x+b_o=0 \qquad (5)$$

w_o→Optimum values of the weight vector, where W={w_1,w_2,…w_n},n is the number of attributes

b_o→Optimum values of the bias and b is scalar. Regard as training sample $\{(x_i, y_i)\}^{N}_{i=1}$, where

x_i→ input pattern and

y_i→ desired output. Pattern represented by the subset

y_i= +1 or -1these pattern representation are linearly separable in "(1)"

4. COMPUTATIONAL ANALYSIS

The proposed methodology is experimented with the dataset namely Cleveland, Hungarian, Switzerland, leukemia, lung, Breast. These datasets are taken from the UCI machine learning repository. It was recorded using Intel® Atom ™ CPU N455@ 1.67GHZ processor and 2GB RAM .the proposed methodology implemented by Mat lab.The computational Analysis result shows cluster center value for each cluster c1,c2 of the high dimensional datasets in Table.1.Various feature selection results shows original feature and selected features in Table.2. The comparison of classification accuracies with different feature selection methods chart in fig .1.

Table.1.cluster centroid value for high dimensional datasets

Data sets	C1 centroid	C2 centroid
Cleveland	1.206147	0.557318
Hungarian	0.311423	0.414436
Switzerland	1.821604	1.787654
Breast	0.682458	0.142903
Leukemia	1.55441	1.722316
Lung	1.619517	2.249127

Table.2. Computational results of various feature selection techniques

Data set	Cleveland		Hungarian		Switzerland		Leukemia		Lung		Breast	
FS method	#O F's	#S F's	#O F's	#S F's	#O F's	#S F's	#O F's	#S F's	#O F's	#S F's	#O F's	#S F's
FRS	13	7	13	8	13	8	7129	82	2000	25	2001	562
FCM-RST	13	6	13	7	13	8	7129	75	2000	21	2001	494
MFCM-RST	13	6	13	6	13	5	7129	71	2000	20	2001	492

#O F's→Original Features, #S F's→ Selected Features

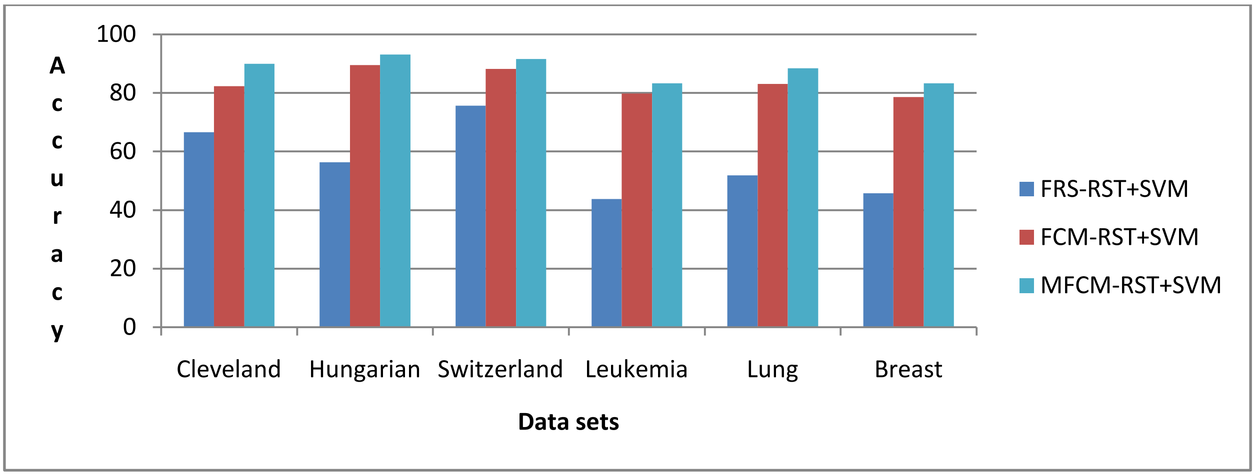

Fig1. The comparison of classification accuracies with different feature selection methods

CONCLUSION

Accuracy is most significant in the field of medical diagnosis to diagnose the patient's disease. Experimental results show that Feature Selection, a Preprocessing technique greatly boost the accuracy of classification. It also concludes that the classifier accuracy has been certainly enhanced by the use of any of Feature selection method than the classifier accuracy without feature selection. For data mining and machine learning problems Feature selection, as an important data preprocessing strategy, has been proven to be effective and efficient in preparing high-dimensional data. The objectives of feature selection include building simpler and more comprehensible models, improving data mining performance, and preparing clean, understandable data. The recent proliferation of big data has presented some substantial challenges and opportunities for feature selection research. In this survey, provide a comprehensive and structured overview of recent advances in feature selection research. Motivated by current challenges and opportunities in the big data age, it revisit feature selection research from a data perspective, and review representative feature selection algorithms for generic data, structured data, heterogeneous data and streaming data. Once the best feature selection method is identified for a particular dataset the same can be used to enhance the classifier accuracy.

REFERENCES

1) K. Jain and R. C. Dubes, Algorithms for Clustering Data. Englewood Cliffs, NJ: Prentice-Hall, 1988.
2) R. O. Duda, P. E. Hart, and D. G. Stork, Pattern Classification and Scene Analysis. New York: Wiley, 1999.
3) J. C. Bezdek, Pattern Recognition With Fuzzy Objective Function Algorithm. New York: Plenum, 1981.
4) R. Krishnapuram and J. M. Keller, “A possibilistic approach to clustering,” IEEE Trans. Fuzzy Syst., vol. 1, no. 2, pp. 98–110, May 1993.
5) N. R. Pal, K. Pal, J. M. Keller, and J. C. Bezdek, “A possibilistic fuzzy c-means clustering algorithm,” IEEE Trans. Fuzzy Syst., vol. 13, no. 4, pp. 517–530, Aug. 2005.
6) F. Masulli and S. Rovetta, “Soft transition from probabilistic to possibilistic fuzzy clustering,” IEEE Trans. Fuzzy Syst., vol. 14, no. 4, pp. 516–527, Aug. 2006.
7) Z. Pawlak, Rough Sets, Theoretical Aspects of Reasoning About Data. Dordrecht, The Netherlands: Kluwer, 1991.
8) P. Lingras and C. West, “Interval set clustering of web users with rough K-means,” J. Intell. Inf. Syst., vol. 23, no. 1, pp. 5–16, 2004.
9) D. Dubois and H. Prade, “Rough fuzzy sets and fuzzy rough sets,” Int. J. Gen. Syst., vol. 17, no. 2/3, pp. 191–209, 1990.
10) S. Mitra, H. Banka, and W. Pedrycz, “Rough–fuzzy collaborative clustering,” IEEE Trans. Syst., Man, Cybern. B, Cybern., vol. 36, no. 4, pp. 795–805, Aug. 2006.
11) A.Pethalakshmi ,A.Banumathi “A novel approach in clustering via rough set.Internatonal journal of Science and Research, Volume 2 Issue 7, pp.139-145,July 2013
12) Y.L.Zhang, N.Guo, H.Du and W.H Li, “Automated defect recognition of C- SAM images in IC packaging using Support Vector Machines,” The International Journal of Advanced Manufacturing Technology 25, 1191 -1196,2005.
13) Karabatak, M., Ince, M.C., “An expert system for detection of breast cancer based on association rules and neural network,” Expert Systems with Applications, Vol.36, pp.3465-3469, 2009.
14) A.Mehmet Fatih, “Support vector machines combined with feature selection for breast cancer diagnosis,” Expert Systems with Applications, Vol. 36, pp.3240-3247,2009.
15) Jollife, I. T. Principal Component Analysis. Springer-Verlag, New York. 1986
16) Eisen MB, Spellman PT, Brown PO and Botstein D: “Cluster analysis and display of genome-wide expression patterns”, Proceedings of the National Academy of Sciences of the United States of America, vol.95, no.25, pp.14863-14868, 1998.
17) Tavazoie S, Hughes JD, Campbell MJ, Cho RJ, Church GM: “Systematic determination of genetic network architecture”, Nature genetics, vol.22, no.3, pp. 281-285, 1999.
18) L. Polkowski, “Rough Sets- Mathematical Foundations”, Physica-Verlag Heidelberg, 2003
19) S. Dudoit, J. Fridlyand and T.P. Speed, “ Comparison of discrimination methods for the classification of tumors using gene expression data” Journal of the American Statistical Association,vol.97,no.457,2002.
20) Q. Hu, Z. Xie, and D. Yu, "Hybrid attribute reduction based on a novel fuzzy-rough model and information granulation," Pattern recognition, vol. 40, pp. 3509-3521, 2007.
21) F. Xu, J. Lei, and L. Wei, "Improved mutual information-based gene selection with fuzzy rough sets," Journal of Computational Information Systems, vol. 7, pp. 3166-3173, 2011.
22) Cornelis, R. Jensen, and G. Hurtado, "Attribute selection with fuzzy decision reducts," Information sciences, vol. 180, pp. 209-224, 2010.
23) R. Jensen and Q. Shen, "New approaches to fuzzyrough feature selection," Fuzzy Systems, IEEE Transactions on, vol. 17, pp. 824-838, 2009.
24) R. Jensen and Q. Shen, "Fuzzy-rough sets assisted attribute selection," Fuzzy Systems, IEEE Transactions on, vol. 15, pp. 73-89, 2007.
25) H. FAN and Y. ZHONG, "A Rough Set Approach to Feature Selection Based on Wasp Swarm Optimization," Journal of Computational Information Systems, vol. 8, pp. 1037-1045, 2012.

