Australian Journal of Basic and Applied Sciences, 10(12) July 2016, Pages: 160-168

AUSTRALIAN JOURNAL OF BASIC AND APPLIED SCIENCES

ISSN:1991-8178 EISSN: 2309-8414
Journal home page: www.ajbasweb.com

G-FeSPO - GSO Based Feature Selection and Parameter Optimization in Fuzzy Supervised Classification

K. Joshil Raj, S. Siva Sathya

Department of Computer Science, Pondicherry University, Pondicherry, India.

Address For Correspondence:
K. Joshil Raj, S. Siva Sathya, Department of Computer Science, Pondicherry University, Pondicherry, India
E-mail: joshlion89@gmail.com

ARTICLE INFO

Article history:
Received 26 April 2016
Accepted 21 July 2016
Published 30 July 2016

Keywords:
Evolutionary algorithm; Group Search Optimizer; Fuzzy Membership Classification; Machine learning; Feature Selection

ABSTRACT

Group Search Optimizer (GSO) is a population based optimization algorithm inspired by animal searching behavior to design optimum searching strategies for solving continuous optimization problems. This paper proposes the use of a hybrid GSO-Fuzzy Supervised classification methodology for feature selection which can select relevant feature subsets and also optimize the parameters of Fuzzy Supervised classifier so as to achieve maximum classification accuracy. GSO has been used to optimize the weight parameters and K-Nearest Neighbours of a novel fuzzy supervised classifier having two membership functions namely Global and Nearest-Neighbour membership functions. The aim is to achieve maximum classification accuracy with optimum values of feature subsets from the dataset by deriving the optimum weights associated with both Global and Nearest-Neighbour membership vectors and the optimum number of K Nearest neighbours of the fuzzy supervised classifier. To test the quality and effectiveness of the proposed methodology, it has been evaluated on standard machine learning medical datasets because they pose real challenges in classification due to their large feature space.

INTRODUCTION

In the field of machine learning, Classification algorithms are designed to learn a function using the training dataset to map an unknown object with large vector of attributes into one of the several categorical classes. Several algorithms have been used in the past for classification namely Neural Network (Tan *et al*, 2006), Support Vector Machine (SVM) (Kotsiantis, 2007), Bayesian Networks (Alexandra *et al*, 2010), Rule based Classifiers (Witten, 2005), Fuzzy k-NN (Keller, 1985), Genetic Algorithm (GA) (Kenneth and Jong, 2006), Ant Colony Optimisation (ACO) (Dorigo and Stützle, 2004), and Particle Swarm Optimization (PSO) (Kennedy and Eberhart, 1995). Of late, Fuzzy Logic (FL) techniques have been used effectively for the tasks of classification. Fuzzy logic poses the ability to mimic the human mind to effectively employ modes of reasoning that are approximate rather than exact (Hofmann, 2005). FL can model nonlinear functions of arbitrary complexity to a desired degree of accuracy. FL is a convenient way to map an input space to an output space. The basis of the fuzzy algorithm (Keller, 1985) is to assign membership as a function of the vector's distance from its nearest neighbors and those neighbors' memberships in the possible classes. A Fuzzy supervised classifier based on the fuzzy membership approach is used in this paper for the purpose of classification. For each test data instance, Two fuzzy membership vectors, Global Membership Vector (GMV) and K-Close Membership Vector (KMV) have been derived using the whole training dataset and nearest neighbor instances respectively. A Class Determinant Function (CDF) is then superimposed on GMV and KMV to get a Weighted Membership Vector (WMV) which represents the degree of belonging of a test data record to a particular class.

To Cite This Article: K. Joshil Raj, S. Siva Sathya., G-FeSPO - GSO Based Feature Selection and Parameter Optimization in Fuzzy Supervised Classification. ***Aust. J. Basic & Appl. Sci.,*** *10(12): 160-168, 2016*

Medical databases pose a great challenge in the classification process as they are characterized by their large feature space. Features may contain false correlations, which hinder the process of classification. Some features may be redundant since the information they add is contained in other features. Feature selection is data preprocessing step which helps in substantial reduction of the input data. It improves classification by searching for the subset of features, which best classifies the data (Srinivas *et al*, 2003). Feature selection techniques do not alter the original representation of the variables, but merely select a subset of them. The objectives of feature selection are manifold, the most important ones being: (a) to avoid over fitting, (b) to provide faster and more effective models, (c) helps in better data understanding of data, (d) reducing measurement and storage requirements, (e) reducing training and testing time and (f) thus improving classifier performance. However, the search for a subset of relevant features introduces an additional layer of complexity in the modeling task.

This paper proposes a method for feature selection on the medical dataset by selecting the most optimal subset of significant features from a complete feature set. Irrelevant and redundant features with poor prediction ability are removed. Optimization of the weights of Global Membership Vector (GMV) and K-Close Membership Vector (KMV) and the number of K-Nearest Neighbours used in the fuzzy classifier has also been done to achieve the maximum classification accuracy. The feature subset selection and the parameter optimization of the Fuzzy Supervised Classifier are accomplished by use of a population based optimization algorithm, Group Search Optimizer (GSO) (He *et al*, 2009). Each member in GSO represents a subset of features and parameter values of W_g (global weight), W_k (k-close weight) and K (number of K-Nearest Neighbours). In each iteration, GSO members evolve to generate better fitness value (classification accuracy) for specific feature subsets and specific values of the parameters. The GSO member with the highest classification accuracy gives the optimal feature subset as well as the optimal values of W_g, W_k, and K. This process significantly enhances the overall performance of dataset classification.

The rest of the paper is organized as follows: A brief introduction to Basic Classification Problem is given in Section 2. Section 3 presents some of the related works. Section 4 introduces the Fuzzy classification System and Section 5 describes the GSO algorithm. The proposed feature selection and parameter optimization of the fuzzy supervised classifier using GSO is described in Section 6 and the experimental analysis is given in Section 7. Finally section 8 concludes the paper.

Basic Classification ProbleM:

Classification or supervised machine learning is a process where the classifier algorithm reasons from externally supplied training instances to produce general hypotheses and makes predictions about future instances. Suppose, each instance i of a given dataset is described by both a vector of n attribute values $x_i=[x_{i1},x_{i2},\ldots,x_{in}]$ and its corresponding class label y_i, which can take any value from a set of values $Y=\{y_1,y_2,\ldots,y_m\}$. Thus, x_{iq} will specify the value of the q-th attribute of i-th instance. The training and test datasets are represented by D_{TRAIN} and D_{TEST} respectively. D_{NN} contains k- Nearest Neighbor records based on the distance between $x\in D_{TEST}$ and $x_i\in D_{TRAIN}$. Therefore, $D_{NN} \subseteq D_{TRAIN}$ and

$$D_{TRAIN} = \{(X_i, Y_i) \mid i = 1,2 \ldots, h\}$$
$$D_{NN} = \{(X_i, Y_j) \mid i = 1,2 \ldots, k\};$$
$$D_{TEST} = \{(X_i) \mid i = 1,2 \ldots, r\}$$

Where h, k, r represents the total number of records in D_{TRAIN}, D_{NN}, D_{TEST} respectively and k may be any arbitrary value between 1 and h. Each test data record from D_{TEST} will be assigned a class label based on the prediction provided by the classifier using the knowledge of the training set D_{TRAIN}.

Related Work:

Various models of Fuzzy classification approaches have been used extensively for the purpose of supervised classification. Nearest neighbor based classifier i.e; k-NN is used in real life applications due to its simplicity and standard performance. Different fuzzy techniques are applied on simple k-NN algorithm to improve its accuracy rate. The first fuzzy nearest neighbour classifier was introduced by Jów´ik in 1983. This classifier is based on a learning scheme of class memberships, provided that each training object is having a membership array which defines its fuzzy membership to each class. Every neighbour uses its membership array for the voting rule which brings the final classification result. In 1985, Keller *et al* proposed a new fuzzy membership based classification method popularly known as FuzzyKNN which uses three different methods for computing the class memberships. The best performing method is considered for each instance x (x is in class i) to compute the K nearest neighbors from the training data. The membership values obtained from the training phase are used for each neighboring instance to vote for every class. The weighted votes, achieved according to the inverse of the distance to the instance to be classified, are added finally. The final class is assigned which has the greatest combined vote. S. Hadjitodorov applied intuitionistic fuzzy set theory to develop the fuzzy nearest neighbour classifier with the non-membership concept in 1995. In this classification technique, the distances, computed by a k-NN classifier are modified using the pattern degrees of memberships and non-membership. In this procedure actual instances get a high degree of membership whereas noisy instances get a high degree of non-membership. A general categorization of the fuzzy nearest neighbour algorithms is shown in the fig.1 based on the taxonomy found in the literature.

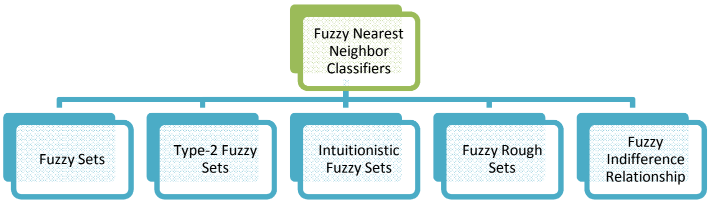


Fig. 1: Taxonomy of different Fuzzy-Nearest Neighbour classification techniques.

José *et al* in 2002 proposed an optimized version fuzzyKNN classification model using genetic algorithm. In 2005, T.D. Pham mentioned in his research work that conventional fuzzy k-NN algorithm depends on the choice of some distance function, which is not based on any principle of optimality. He introduced the kriging computational scheme, known as the best linear unbiased estimator, for determining optimal weights to be combined with different fuzzy member-ship grades for classification in the fuzzy k-NN approach. This linear estimation aims to minimize the error variance estimating the weighted linear combinations of the available data and imposes a condition that the mean error is equal to zero. This optimally weighted fuzzy k-NN algorithm is effectively applied on the microarray-based cancer data which performs better than both k-NN and conventional fuzzy k-NN algorithms in terms of predicted accuracy. Recently in 2011, Feras Al-Obeidat *et al* proposed a new methodology named as PSOPRO based on fuzzy indifference relationship and PSO (Particle Swarm Optimization) algorithm to train the Multi-Criteria Decision Aid (MCDA) method called PROAFTN. PSOPRO employs PSO for training and improving the efficiency of the PROAFTN classifier. During the learning stage, PSO uses training samples to induce the best PROAFTN parameters in the form of prototypes. Then, these prototypes, representing the classification model, are used for assigning unknown samples. In 2012, M. Sarkar proposed a fuzzy-rough uncertainty based classification method which is more efficient than the conventional K-nearest neighbour algorithm. This classifier incorporates the lower and upper approximations of the memberships to the decision rule, and deals with both fuzzy uncertainties and rough uncertainties. The classifier does not require knowing the K value for the classification rule and in this case the generated class confidence values known as fuzzy-rough ownership values do not necessarily sum up to one. Later in 2012, Emel Kizilkaya Aydogan *et al*. introduced a fuzzy rule-based (FRBCSs) hybrid heuristic approach (hGA), based on genetic algorithm (GA) and integer-programming formulation (IPF), to solve high dimensional classification problems in linguistic fuzzy rule-based classification systems. Earlier in 2006, He *et al*. successfully modelled GSOANN classification method, in which, Group Search Optimizer is used to optimize the three layers feed forward Artificial Neural Network. Support vector machines and neural network were used for feature selection and classification by Srinivas and Sung et.al in 2003. Features were deleted one at time and then classification accuracy was calculated. Reducing the number of input features had statistically significant impact on the accuracy of classification. Each reduced feature set was also tested on SVM and Neural Networks to rank the importance of the input features. The reduced feature set that yielded the best detection rate in the experiments was considered to be the important features subset.

GSO AlgorithM:

Group search optimizer (GSO) [11] is a population based optimization algorithm, inspired by animal searching (foraging) behavior. GSO Optimization employs the producer–scrounger (PS) model and the animal scanning mechanism. The population of the GSO is called a group, where each individual is called a member. A group consists of three types of members: producers, scroungers and rangers. Producers perform producing strategy in the way of animal scanning mechanism; scroungers perform scrounging strategy by joining resources uncovered by others; and rangers search for the randomly distributed resources by random walks. In each generation, the best fit member is treated as the producer, and a number of members except the producer in the group are selected as the scroungers, while the remaining members are regarded as the rangers.

In an n-dimensional search space, the i^{th} member at the k^{th} searching bout (iteration) has a position, $X_i^k \in \mathbb{R}^n$ and a head angle $\Phi_i^k = (\Phi_{i1}^k, \Phi_{i2}^k, \ldots\ldots \Phi_{i(n-1)}^k) \in \mathbb{R}^{n-1}$. The search direction associated with the i^{th} member, which is represented as a unit vector $D_i^k(\Phi_i^k) = (d_{i1}^k, d_{i2}^k, \ldots\ldots d_{in}^k) \in \mathbb{R}^n$ as presented in equation(2).

$$\begin{cases} d_{i_1}^k = \prod_{q=1}^{n-1} \cos(\Phi_{i_q}^k) \\ d_{i_j}^k = \sin(\Phi_{i_{(j-1)}}^k).\prod_{q=j}^{n-1} \cos(\Phi_{i_q}^k)\ (j = 2,\dots,n-1) \\ d_{i_n}^k = \sin\left(\Phi_{i_{(n-1)}}^k\right) \end{cases} \tag{1}$$

The scanning field of vision is an n-dimensional space, which is characterized by maximum pursuit angle $\Theta_{max} \in \mathbb{R}^1$ and maximum pursuit distance $l_{max} \in \mathbb{R}^1$. In GSO, the producer X_p^k randomly samples three points in the scanning field: one point at zero degree, one point in the right hand side hypercube, and one point in the left hand side hypercube, as listed in equation (2).

$$\begin{cases} X_z = X_p^k + r_1 l_{max} D_p^k(\Phi^k) \\ X_r = X_p^k + r_1 l_{max} D_p^k(\Phi^k + r_2\, \Theta_{max}/2) \\ X_l = X_p^k + r_1 l_{max} D_p^k(\Phi^k - r_2\, \Theta_{max}/2) \end{cases} \tag{2}$$

Where, $r_1 \in R^1$ is a normally distributed random number with mean 0 and standard deviation 1 and $r_2 \in R^n - 1$ is a uniformly distributed random sequence in the range (0, 1). In GSO, the producer X_p^k scans at zero degree and laterally by randomly sampling three points in the search space to find the best resource at each iteration. If the producer cannot find a better search position after 'a' iterations, it will turn its head back to zero degree. Scroungers follow the producer adopting a random walk towards it.

$$X_i^{k+1} = X_i^k + r \circ (X_p^k - X_i^k) \tag{3}$$

where $r \in \mathbb{R}^n$ is a uniform random sequence in the range (0,1). Operator "o" is the Hadamard product or the Schur product, which calculates the entry wise product of the two vectors. If a better position than the current producer is found by any of the scroungers then in the next searching bout it will switch to be a producer. This switching mechanism helps the group members to escape from local minima in the previous search bouts. Dispersed animals may adopt ranging behaviour to explore and colonize new habitats. In each generation, rangers move to the new point based on a random head angle and a random distance using the following equation:

$$\begin{cases} l_i = a.r_1.l_{max} \\ X_i^{k+1} = X_i^k + l_i.d_i^k.(\Phi^{k+1}) \\ \Phi^{k+1} = \Phi^k + r_2\alpha_{max} \end{cases} \tag{4}$$

Here, $r_1 \in \mathbb{R}^1$ is a normally distributed number with mean 0 and standard deviation 1 and $r_2 \in \mathbb{R}^{n-1}$ is a uniformly distributed random sequence in the range (0, 1).

Fuzzy Classification System:

The advantage provided by fuzzy sets is that the degree of membership of an object in a set can be specified, rather than just the binary. Fuzzy technique specifies to what degree the object belongs to each class, which is given by a characteristic function u: U → [0, 1]. The fuzzy classification approaches also restricts the criteria that the sum of the membership degree to every class of each instance must be 1 for mathematical tractability. Therefore, if $u_i \in [0,1]$is the membership of ith class of a particular instance then,

$$\sum_{i=1}^{c} u_i = 1 \text{For i=1, 2…, m} \tag{5}$$

Where 'm' is the number of classes present in the classification model.

KNN (K-Nearest Neighbors) consists of the identification of groups of individuals with similar features and posterior grouping. Let the training data (D_{TRAIN}), test data (D_{TEST}), and nearest neighbor data (D_{NN}) sets be defined as described in Section-2. The adjusted weight,$\eta_i(x)$, for each distance d_i (Euclidean Distance) between test data x to training data x_i is formulated below in equation(6). Here d_i is the Euclidean distance between test data and ith training data. The fuzzy membership vectors (Global and k-close) for each test data are using these adjusted weights.

$$\eta_i(x) = 1/(1+d_i) \qquad \text{Where, } d_i = \sqrt{\sum_{q=1}^{n}(x_q - x_{iq})} \tag{6}$$

For each test data we define the Global Membership Vector and K-Close Membership Vector as GMV = (GM_1, GM_2 …….GM_m) and KMV = (KM_1, KM_2 ……..KM_m) respectively. Here, m is the number of classes present in training dataset. These fuzzy membership vectors are evaluated using equations (7) and (8) given below.

$$GMj = \frac{\text{Total adjusted weight of } j\text{th class label records in } D_{TRAIN}}{\text{Total adjusted weight of all class label records in } D_{TRAIN}} = \frac{\sum_{(x_i, y_j) \in D_{TRAIN}} \eta_i(x)}{\sum_{i=1}^{h} \eta_i(x)} \quad (7)$$

$$KMj = \frac{\text{Total adjusted weight of } j\text{th class label records in } D_{NN}}{\text{Total adjusted weight of all class label records in } D_{NN}} = \frac{\sum_{(x_i, y_j) \in D_{NN}} \eta_i(x)}{\sum_{i=1}^{k} \eta_i(x)} \quad (8)$$

A Weighted Membership Vector (WMV) is found after deriving the GMV and KMV vectors for the test data $x \in D_{TEST}$ using a special Class Determinant Function (CDF) given in equation (9). It indicates a Determinant Function Vector, DFV = (F_1, F_2……, F_m).

$$< DFV >= w_g \times < GMV > + w_k \times < KMV > \quad (9)$$

Where, $F_j = w_g \times GM_j + w_k \times KM_j$. Here, '×' represents the scalar multiplication with vectors and '+' denotes the vector addition. $W_g \in (0, 1)$ and $w_k \in (0, 1)$ are the global weight and k-close weight respectively. These two variables define the weightage given to global membership vector and k-close membership vector respectively. The Weighted Membership Vector (WMV) = (g_1, g_2…., g_m) may be achieved by normalizing the values in Determinant Function Vector, DFV.

The last step is to assign each test data record, $a \in D_{TEST}$ to the right class y_j, where $j \in \{1, 2 . . . , m\}$. The actual class for the test data record, $x \in D_{TEST}$ is evaluated by applying the following rule i.e; equation (10).

$$X \in y_j \text{ if and only if } \max(g_1, g_2 \ldots, g_m) = g_j \quad (10)$$

Proposed GSO based Feature Selection and Parameter Optimization:

Feature selection is one of the major tasks in a classification process especially in medical datasets that have a large number of features. The main objective of feature selection is to retain the optimum salient characteristics necessary for the classification process even while reducing the dimensionality of the measurement space. It is important that the extracted features are relevant to the particular task at hand. Errors or noise may also get introduced in the process of extracting feature subsets. The problem of feature selection has two aspects: the selection of an optimal subset from the available features and the formulation of a suitable criterion to evaluate the goodness of a set of features. In some cases there are mathematical tools that help in feature selection. In other cases, simulation may aid in the choice of appropriate features. Majority of learning algorithms perform well on domains with relevant information. They degrade in adverse situations like: data with high noise content, small sample sizes relative to number of features, irrelevant or redundant information and non-linearity [22]. Features may contain false correlations, which hinder the process of detecting intrusions. Improved performance may be achieved by discarding such noisy, irrelevant and redundant information. Feature selection is selection of a subset of features that describe the hypothesis at least as well as the original set. Feature selection improves classification by searching for the subset of features, which best classifies the training data. The benefits of feature selection thus include a reduction in the amount of data needed to achieve learning, improved predictive accuracy, more compact and easily understood knowledge base and reduced execution time.

1. GSO based Feature Selection and Parameter Optimisation. Each member in GSO represents a set of features and the values of W_g (global weight), W_k (k-close weight) and K (number of Nearest-Neighbours). Each GSO member will have values in '3 + n' dimensions. K, W_g and W_k are the 3 Fuzzy supervised classifier parameters that are to be optimized and 'n' will represent the number of features. In each generation GSO members will evolve with producing, scrounging and ranging mechanism. The fitness function of each member is the classification accuracy found with the corresponding feature subsets being used from the data and the using the parameter values associated with that GSO member. After termination of GSO procedure, the producer (the best fit member) will indicate the optimal values of the KNN, W_g and W_k which produces the optimal classification accuracy.

The individual GSO member representation is as shown below:

K	W_g	W_k	A_1	A_2	…..	…..	A_n

Where K, W_g, and W_k are the parameters of Fuzzy supervised classifier that is to be optimized and variable (A_1, A_2……A_n) are the features of the dataset. If the value of the variable (A_1, A_2…….A_n) is less than or equal to 0.5, then its corresponding feature is not chosen. Conversely, if the value of a variable is greater than 0.5, then its corresponding feature is chosen for making the feature subset. The searching range of parameter K is between 1 and $\sqrt{(2h+1)}$ where, 'h' is the total number of records in training set. W_g and W_k are any real valued number between 0 and 1.

2. Fitness Function. Classification accuracy and the number of selected features are the two criteria used to design a fitness function. Thus, for the GSO member with high classification accuracy and a small number of features produce a high fitness value. The fitness of i_{th} member is using equation (11) given below:

$$fitness_i = W_A \times acc_i + W_F \times \left[1 - \frac{(\sum_{j=1}^{n_F} f_j)}{n_F}\right] \quad (11)$$

Where w_A is the weight for the Fuzzy Supervised classification accuracy. w_A is adjusted to 95%,

acc_i is the Fuzzy Supervised classification accuracy.

w_F is the weight for the number of selected features. w_F is set to 5%.

f_i is the value of feature mask - '1' represents that feature i is selected and '0' represents that feature 'i' is not selected.

n_F is the total number of features.

The flow chart for the GSO based feature selection and parameter optimisation is as shown in Figure 1:

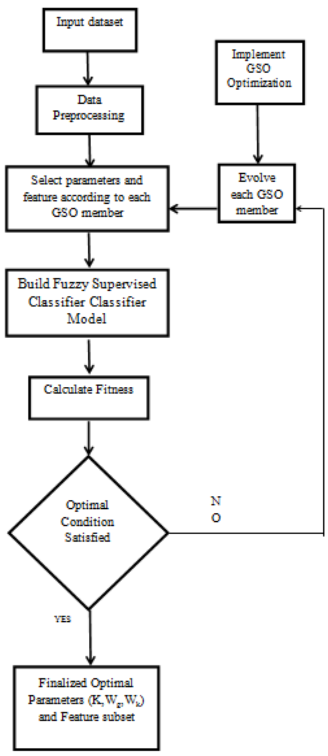
Experimental Analysis:

The proposed methodology is implemented in Java using the open source java packages of WEKA 3.7.9 machine learning tool developed by Waikato University. Performance is evaluated on the some of the medical datasets present in UCI benchmark data sets. The datasets are available at the University of California at Irvine (UCI) Machine Learning repository database. In addition to these datasets, three microarray datasets from the Kent Ridge Biomedical datasets repository and the LibSVM repository have also been considered for validating the results.

The Colon Cancer dataset was taken from the Kent Ridge Biomedical datasets repository and comprises of 62 samples of cell lines retrieved from colon cancer patients. 40 samples are of tumor patients and 22 samples are normal samples. The Leukemia dataset consists of the expression of 7129 instances (genes) retrieved from 72 samples. 22 of these samples are of Acute Myeloid Leukemia (AML) class and 47 belong to the Acute Lymphoblastic Leukemia (ALL) class. The Duke breast cancer dataset consists of 44 samples of which 22 belong to one class and 22 to the other class. The dimensions of the datasets are tabulated in Table 1:

Table 1: Bio-medical datasets specifications

BIO-MEDICAL DATASETS	NO OF FEATURES	NO OF INSTANCES	NO OF CLASSES
Leukemia	7129	72	2
Colon Cancer	2000	62	2
Duke Breast Cancer	7129	44	2



1. Experimental Settings. The parameter values used in GSO are summarized as follows. The initial population of GSO is generated uniformly at random in the search space. The initial head angle Φ^0 of each individual is set to be ($\pi/4 \ldots \pi/4$). The constant 'a' is chosen as round ($\sqrt{m+1}$) where m is the dimension of the search space i.e. number of distinct classes in training dataset. The maximum pursuit angle Θ_{max} is π/a^2. The maximum turning angle α_{max} is set to be $\Theta_{max}/2$. The maximum pursuit distance l_{max} is calculated using equation (12) given below:

$$l_{max} = \| U - L \| = \sum_{i=1}^{m} (U_i - L_i)^2 \quad (12)$$

In the proposed method U_i=1 and L_i=0 for i= {1, 2…... m}. In each generation, one best fit member is treated as the producer, 20% of the total GSO population are chosen randomly as rangers and remaining members will perform scrounge operation. The population size is taken as 30 and the number of iteration as 80 in the proposed method.

2. Performance Evaluations. To evaluate the performance of the proposed Feature selection and parameter optimization of Fuzzy supervised classifier using GSO, 10-fold cross validation classification accuracy is generated for each dataset using the complete dataset and the dataset with selected feature subsets and the optimised parameter values. In this process, each time a different set is used for testing (i.e., 1/10 of data) and the remaining data (i.e., 9/10 of data) are used for training. This procedure is repeated until each partition of the dataset is tested. Table 2 and Table 3 gives the comparison of the classification accuracy achieved using the complete dataset and the reduced feature dataset derived from the UCI dataset repository. Table 3 gives the results for the bio-medical datasets from the UCI dataset repository. The optimised values of the parameters used in the Fuzzy supervised classifier are also given in these tables.

Table 2: Comparative classification accuracy (%) results for UCI datasets

DATASETS	COMPLETE DATASET		REDUCED DATASET				
	FEATURES	ACCURACY (%)	SELECTED FEATURES	OPTIMIZED PARAMETERS			ACCURACY (%)
				W_g	W_k	KNN	
Australian	14	86.38	14	0.857	0.999	9	86.67
Corral	6	92.19	5	1	0.773	3	100
Flare	10	83.11	6	0.773	0.491	1	83.48
Glass	9	79.44	9	1	1	3	79.90
Iris	4	94.67	3	0.048	1.00	3	95.33
Vehicle	18	73.64	17	1	1	4	73.75
Vote	16	93.56	10	0.874	0.325	3	95.63
Soybean-Large	35	91.07	34	0.722	0.96	1	91.80

Table 3: Comparative classification accuracy (%) results for UCI bio-medical datasets

DATASETS	COMPLETE DATASET		REDUCED DATASET				
	FEATURES	ACCURACY (%)	SELECTED FEATURES	OPTIMIZED PARAMETERS			ACCURACY (%)
				W_g	W_k	KNN	
Breast Cancer	9	97.42	9	1	1	6	97.13
Cleveland Heart	13	84.46	13	0.999	0.987	6	85.81
Diabetes	8	77.86	8	0.732	0.965	17	77.86
Heart	13	84.81	12	0.554	0.621	5	85.18
Hepatitis	19	87.74	19	0.465	0.939	6	88.38
lymphography	18	85.14	12	0.00	1	2	87.83

Conclusion:

In this paper, parameter optimization of fuzzy supervised classification and feature selection is proposed using Group Search Optimization technique, a population based optimization algorithm developed from the group search behaviour of animals in search of food. Global and K-close weights associated with Global and the Nearest-Neighbour membership vectors respectively and the number of nearest neighbours employed in the fuzzy supervised classification methodology are optimized so as to generate the highest classification accuracy. Alongside, the dataset is also reduced by removing redundant and irrelevant features using the GSO algorithm. The advantage is that both feature selection and parameter optimization are done using the same GSO in a single run of the algorithm thereby reducing the burden of the classifier. This methodology was tested on the medical datasets as well as some standard datasets which shows that classification accuracy has significantly increased, thereby proving it to be an effective tool for feature reduction and optimization followed by classification.

REFERENCES

Pang-Ning Tan, Michael Steinbach and Vipin Kumar, 2006. Introduction to Data Mining. Pearson Education, Inc.

Kotsiantis, S.B., 2007. “Supervised Machine Learning: A Review of Classification Techniques,” Informatica, 31: 249-268.

Alexandra, M. Carvalho, Teemu Roos, Arlindo L. Oliveira, 2011. "Discriminative Learning of Bayesian Networks via Factorized Conditional Log-Likelihood," Journal of Machine Learning Research, 12: 2181-2210.

Witten, H., 2005. Data Mining: Practical Machine Learning Tools and Techniques, Morgan Kaufmann Series in Data Management Systems.

Keller, J.M., M.R. Gray, J.R.. Givens, J.A. A Fuzzy K-Nearest Neighbor Algorithm, 1985. IEEE Transactions on Systems, Man, and Cybernetics, v. SMC-15, n. 4: 258-260.

Kenneth A., 2006. De Jong, Evolutionary Computation A Unified Approach. The MIT Press, Cambridge, England.

Marco Dorigo and Thomas Stützle, 2004. Ant Colony Optimization, The MIT Press Cambridge, England.

Kennedy, J. and R. Eberhart, 1995. "Particle swarm optimization," Proc. IEEE Int. Conf. Neural Networks, pp: 1942-1948.

Hofmann, H.D., 2005. "Application of Intelligent Measurements with Metrical Image Processing for Quality Control", resented at the 5th International Conference, PEDAC' 92, Alexandria, EGYPT.

Mukkamala Srinivas, Sung Andrew H, Abraham Ajith, 2003. Intrusion detection using ensemble of soft computing paradigms. In: Third international conference on intelligent systems design and applications. Intelligent systems design and applications, advances in soft computing. Germany: Springer Verlag; pp: 239e48.

He, S., Q.H. Wu and J.R. Saunders, 2009. “Group Search Optimizer: An Optimization Algorithm Inspired by Animal Searching Behavior,” IEEE Transactions On Evolutionary Computation, 13(5): 973-990.

Jów´ik, A., 1983. “A learning scheme for a fuzzy k-NN rule”, Pattern Recognition Letters, 1: 287-289.

Hadjitodorov, S., 1995. “An intuitionistic fuzzy sets application to the k-NN method,” Notes on Intuitionistic Fuzzy Sets., 1: 66-69.

Jose, L., A. Rosa and Nelson F.F. Ebecken, 2003. “Data Mining for Data Classification Based on the KNN-Fuzzy Method Supported by Genetic Algorithm,” VECPAR 2002, LNCS 2565, pp: 126-133.

Pham, T.D., 2005. “An optimally wieghted fuzzy k-NN algorithm,” in: Proceedings of the Third International Conference on Advances in Pattern Recognition, Part I (ICAPR 2005), Bath, UK, pp: 239-247.

Feras Al-Obeidat, Nabil Belacel, Juan A. Carretero, Prabhat Mahanti, 2011. “An evolutionary framework using particle swarm optimization for classification method PROAFTN,” Applied Soft Computing, 11: 4971-4980.

Sarkar, M., 2012. “Fuzzy-rough nearest neighbor algorithms in classification,” Fuzzy Sets and Systems, 158: 2134-2152.

Emel Kızılkaya Aydogan, Ismail Karaoglan and Panos M. Pardalos, 2012. “hGA: Hybrid genetic algorithm in fuzzy rule-based classification systems for high-dimensional problems,” Applied Soft Computing, 12: 800-806.

He, S., Q.H. Wu and J.R. Saunders, 2006. "A Group Search Optimizer for Neural Network Training," Computational Science and Its Applications - ICCSA 2006 Lecture Notes in Computer Science., 3982: 934-943.

Mukkamala Srinivas, H. Sung Andrew, 2003. Identifying Important Features for Intrusion Detection Using Support Vector Machines and Neural Networks. Proceedings of the Symposium on Applications and the Internet (SAINT’03).

Chris Ding and Hanchuan Peng, 2005. Minimum redundancy feature selection from microarray gene expression data. journal of bioinformatics and computational biology, 3(2): 185-205.

Kumar, R., V.K. Jayaraman, B.D. Kulkarni, 2005. An SVM Classifier Incorporating Simultaneous Noise Reduction and Feature Selection: Illustrative Case Examples. J. Pattern Recognition., 38: 41-49.

WEKA software, Machine Learning, The University of Waikato, Hamilton,New Zealand. http://www.cs.waikato.ac.nz/ml/weka/

UCI Machine Learning Repository, http://mlearn.ics.uci.edu/MLRepository.html.

Kent ridge bio-medical dataset, URL: http://datam.i2r.astar.edu.sg/datasets/krbd/

Chang, C.-C. and C.-J. Lin, 2011. “LIBSVM: A library for support vector machines,” ACM Transactions on Intelligent Systems and Technology, 2(27): 1-27.

Alon, U., N. Barkai, D.A. Notterman, K. Gish, S. Ybarra, D. Mack and A.J. Levine, 1999. “Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays,” Proceedings of the National Academy of Sciences, 96(12): 6745-6750.

Golub, T.R., D.K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J.P. Mesirov, H. Coller, M.L. Loh, J.R. Downing, M.A. Caligiuri, C.D. Bloomfield and E.S. Lander, 1999. "Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring, SCIENCE, www.sciencemag.org, 286: 531-537.