Chapter 4

Improving the Initialization

4.1 Introduction

In the literature many authors have suggested different techniques for initialization of PSO. Qiang *et al.* [72] suggested chaotic initialization which uses something similar to pheromone used in ACO. Ruksaphil *et al.* [73] proposed minimax initialization which minimizes the maximum error. Paolo *et al.* [74] suggested alternate initialization technique of log, normal and lognormal distribution which replaces uniform distribution, Babu *et al.* [75] suggested two stage initialization, in which first stage is about selecting best strings, by evaluating these strings repeatedly with equal number of population size and in second stage these best strings get combined, and forms the new population, which is used for further operations. Guo *et al.* [76] suggested a re-initialization technique, which is based on estimation of the varieties and activities of the particles. In this method, group of particles which satisfies re-initialization pre-conditions, will be used for initialization, facilitating balance global search capabilities.

The above initialization methods are for general PSO algorithm; these are not specifically for feature selection. In the case of feature selection, an initialization method should ensure that distribution of strings in the population is uniform as far as a number of features selected are concerned. In this chapter a new initialization method based on this idea is developed and tested for feature selection problem.

4.2 Proposed Distributed Sampled (DS) Initialization

In this work, a new initialization technique, has been proposed and has been implemented for PSO and IDS techniques, discussed in the previous section for feature selection. This initialization attempts to explore the search space in a uniform fashion. The proposed initialization techniques distributes the search space in three parts and then takes candidate solutions (samples) randomly from each of three search spaces. Hence, this technique is referred to as Distributed Sample Initialization (DS - initialization) in this paper, to differentiate it from random initialization (RANDOM initialization) used conventionally. This method of initialization takes advantage of population of solutions. This method initializes population with a mix of three types of solution. First One-third of the population contains solutions with less number of features, second one-third part contains solutions with medium number of feature and remaining one-third part of the population contains solutions with large number of features. Presence of all the three types of solutions in the population ensures that algorithm captures suitable reducts quickly.

The procedure of the method has been presented in the Table 4.1 and contrast of the same has been shown with respect to conventional random initialization method. Here, it can be seen that if total population is P, then one-third of the solution (0 to $\frac{1}{3}$P) are biased towards selection of less number of feature due to the probability p_1, which is less than p_2 and p_3. Similarly, second one-third part of the solution (($\frac{1}{3}$P+1) to $\frac{2}{3}$P) are having average number of features because $p_2 < p_3$. Last one-third part of the population (($\frac{2}{3}$P+1) to P) are having higher number of features. Whereas in random initialization the above is not ensured as practical random number generators are biased towards mean.

4.2.1 PSO with DS Initialization: PSO-DS

The PSO algorithm initialized with DS-initialization is given in *Algorithm 4*. In the present work, PSO of reference [4] has been implemented with the DS-initialization and results have compared with that of RANDOM-initialization.

Algorithm 4 DS-initialization for PSO (PSO-DS)

p_1, p_2, p_3, probabilities between (0,1);

For first one third solutions $i = 1$ to $i/3$

if $rand < p_1$ **then**

  $x_{i,j} = 0;$

else

  $x_{i,j} = 1;$

end if

For second one third solutions $i = (i/3 + 1)$ to $2i/3$, initialize X_i

if $rand < p_2$ **then**

  $x_{i,j} = 0;$

else

  $x_{i,j} = 1;$

end if

For last one third solutions $i = (2i/3 + 1)$ to i, initialize X_i

if $rand < p_3$ **then**

  $x_{i,j} = 0;$

else

  $x_{i,j} = 1;$

end if

Use Algorithm 1 for PSO with $x_{i,j}$; $i = 1$ to P; $j = 1$ to D

Table 4.1: Initialization of solutions for random and proposed DS-initialization method

<table>
<tr><th colspan="3">Proposed DS-Initialization</th><th colspan="3">Random Initialization</th></tr>
<tr><th>Part of population</th><th>For every solutions</th><th>Value assigned</th><th>For the whole population</th><th>For every solutions</th><th>Value assigned</th></tr>
<tr><td rowspan="2">0 to $\frac{1}{3}$P</td><td>If $rand < p_1$</td><td>0</td><td rowspan="6">P</td><td rowspan="3">If $rand < p_i$</td><td rowspan="3">0</td></tr>
<tr><td>otherwise</td><td>1</td></tr>
<tr><td rowspan="2">($\frac{1}{3}$P+1) to $\frac{2}{3}$P</td><td>If $rand < p_2$</td><td>0</td></tr>
<tr><td>otherwise</td><td>1</td><td rowspan="3">otherwise</td><td rowspan="3">1</td></tr>
<tr><td rowspan="2">($\frac{2}{3}$P+1) to P</td><td>If $rand < p_3$</td><td>0</td></tr>
<tr><td>otherwise</td><td>1</td></tr>
<tr><td colspan="3">here, $0 < p_1 < p_2 < p_3 < 1$</td><td colspan="3">here, $0 < p_i < 1$</td></tr>
</table>

4.2.2 IDS with DS Initialization: IDS-DS

The IDS algorithm initialized with DS-initialization is given in *Algorithm 5.* In the present work, IDS of reference [29] has been implemented with the DS-initialization and results have compared with that of RANDOM-initialization.

4.3 PSO and IDS with RST measure

In this paper, Rough Set Theory based Feature Selection (RST-FS) method has been implemented. In this method, objective is to select an optimal set of reduced features (reduct) from the set of unreduced features. The dependency of this reduced feature on the corresponding class is maximized in terms of rough dependency measure, $\gamma_P(Q)$. In this work, swarm intelligence optimization methods (PSO and IDS) have been used to find optimal reduced set of features (reduct) using a population of candidate solutions.

A candidate solution in a given population, consists of string of 0's and 1's. Sum of number of 0's and 1's is equal to number of unreduced features in a given dataset. Index of each '1' is the index of feature selected, in the candidate set of reduced feature. Similarly index of each '0' represents index of feature which is not selected in the above set of reduced features. Fitness function of each solutions is calculated in terms of Rough Dependency Measure, $\gamma_P(Q)$.

Algorithm 5 DS-initialization for IDS (IDS-DS)

p_1, p_2, p_3, probabilities between (0,1);

For first one third solutions $i = 1$ to $i/3$, initialize X_i

if $rand < p_1$ **then**

  $x_{i,j} = 0;$

else

  $x_{i,j} = 1;$

end if

For second one third solutions $i = (i/3 + 1)$ to $2i/3$, initialize X_i

if $rand < p_2$ **then**

  $x_{i,j} = 0;$

else

  $x_{i,j} = 1;$

end if

For last one third solutions $i = (2i/3 + 1)$ to i, initialize X_i

if $rand < p_3$ **then**

  $x_{i,j} = 0;$

else

  $x_{i,j} = 1;$

end if

Use Algorithm 3 for IDS with $x_{i,j}$; $i = 1$ to P; $j = 1$ to D

4.3.1 Experiments, Results and Discussions

We have used MATLAB version R2013a for writing the program and core i3 processor 1.7 GHz for running the program. Parameters related to experiment are discussed as follows.

Dataset

All the benchmark dataset are taken from the UCI data repository of machine learning. [77]. Characteristics of the datasets chosen for the present study are described in Table 4.2. Since all the datasets chosen for this work used in this work are having continuous values, discretization (i.e. converting continuous values to nominal values) of all the datasets has been performed in this case when RST-FS method is being implemented.

Table 4.2: Description of benchmark datasets

S. No.	Dataset	No. of Objects	No. of Features
1	Cleveland	303	13
2	Ecoli	336	7
3	Glass	214	9
4	Ionosphere	351	34
5	Lung	32	56
6	Soybean small	47	35
7	Wine	178	13
8	LSVT	126	310

Parameter Setting

Using trial and error on various dataset we came to the conclusion about the parameters that will be suitable for our experiments. In Table 4.1, for random initialization, we took $p_i = 0.5$, and while doing proposed DS-initialization, we took $p_1 = 0.1$, $p_2 = 0.5$, and $p_3 = 0.9$ after trying many combinations of these probabilities. In Equation (1.1), C_1 and C_2 will be having their usual value 2. In PSO the inertia weight decreases from 1.4 to 0.4 using following equation, generation by generation [4].

$$weight = (weight - 0.4) * \left(\frac{Max\_iter - Current\_iter}{Max\_iter}\right) + 0.4 \quad (4.1)$$

Parameters in IDS are set as follows. $C_w = 0.1$, $C_p = 0.4$, $C_g = 0.9$.

We executed programs corresponding to all the methods with 100 generations, and with population size of 100. Execution of program was done a 12 times and the all these runs were averaged.

Fitness value

The expression of fitness value uses the expression of rough dependency measure as first term. In the expression of fitness value, second term is introduced to evolve minimal reduct. The weightage of rough dependency measure is higher as compared to second term [4]. This is done because reducts of different sizes may have value of dependency measure as 1. Thus the expression for the fitness value is as follows.

$$Fitness\_Value = 0.9 * \gamma_P(Q) + 0.1 * \left(\frac{|C| - |R|}{|C|}\right) \quad (4.2)$$

Here $\gamma_P(Q)$ is the fuzzy rough dependency measure of selected feature set, R, relative to decision feature D. C is the total number of features in the dataset.

Statistical Analysis

In order to validate results, statistical t-tests (denoted 'S'), Wilcoxon (denoted 'WT'), and Friedman ranking are performed for both classification accuracy and for the reduct, with respect to DS initialization of PSO and IDS methods. Statistical analysis establishes that the results found are not by chance. The t-test is a parametric test based on the assumption that the subject data groups under comparison are drawn from the normal distribution. In general cases, the normality of data is assumed rather than verified and therefore, the validity of t-test under such circumstances is not reliable. In view of this, non-parametric tests such as the Wilcoxon and Friedman tests are used. Therefore, in this chapter results have also been validated using Wilcoxon and Friedman tests. In t-tests and Wilcoxon test, significance value of 0.05 is taken. The symbol "*" denotes that the proposed methods perform worse than the indicated method, "-" denotes that the proposed methods perform equally well as compared to the indicated method and "v" denotes that the proposed methods performs better than the indicated method. For example in Table 4.8, *Lung* dataset, the symbol "v" marked against IDS-RANDOM indicates that IDS-DS performs better than IDS-RANDOM. Similarly, the symbol "-" marked against PSO-RANDOM indicates that performance of IDS-DS is equally good to that of PSO-RANDOM. Thus, more number of "v" or "-" indicates that IDS-DS is either better or equally good as compared to other methods given in the Table 4.9.

Results and Discussions

Randomly initialized and DS-initialized PSO and IDS have been executed for each of the benchmark datasets using lower approximation based rough sets. Classification accuracies are computed using J48, JRip and PART classifier in terms of their best, mean and standard deviation values. It is observed from the Tables 4.3 and 4.4 that rough set methodology facilitates the drastic reduction in the size of the feature subset. Here PSO and IDS with random initialization and proposed DS initialization are used to perform the feature selection task. The significance of DS initialization is visible in the reasonably large dataset like *Soybean small*, *Lung* and *LSVT*. For example in case of *LSVT*, PSO-DS gives size of reduct as 14 as compared to 113 given by PSO-random and 119 given by IDS-random, where the unreduced feature size is 310. Similar is the case of *Lung* dataset. IDS-DS also yields similar results. Feature sets have high rough dependency measure even with this reduced reduct size. However, in smaller datasets all the methods provide stable sized reducts, i.e. s.d = 0.

PSO-DS and IDS-DS are better as compared to random PSO and random IDS respectively. Further, effect of optimized reducts on classification accuracy has been investigated.

Through this investigation it has been demonstrated that in most of the cases classification accuracy is acceptable. Table 4.3 show results of t-test (denoted 'S') and Wilcoxon test (denoted 'WT') for PSO-DS method in terms of classification accuracy. In this table, statistical significance of PSO-DS as compared to PSO-random and IDS-random has been tabulated. From these values it is evident that the proposed PSO-DS is always comparable in terms of statistical significance in at least one of the three classifiers used in this work.

Table 4.3: PSO-DS with RST measure: Comparison of reduct size and classification accuracy with statistical t-test, and Wilcoxon test. [Classification accuracy is denoted as 'CA'. Statistical t-test is denoted as 'S'. Wilcoxon test is denoted as 'WT'. Mean Subset Size is denoted as 'MSS'.]

Dataset (Total features)	Feature Selection Method	Feature Subset size				Classification Accuracy (CA)												Best result reported in literature	
						Classifier : J48				Classifier : JRip				Classifier : PART					
		Best	Mean(s.d.)	S	WT	Best	Mean(s.d.)	S	WT	Best	Mean(s.d.)	S	WT	Best	Mean(s.d.)	S	WT	MSS	CA
Cleveland(13)	PSO-RANDOM	3	3(0)	-	-	55.44	51.89(3.067)	-	-	**56.1**	53.18(1.032)	-	-	54.78	50.84(2.976)	-	-	7.81 [30]	52.6 [30]
	IDS-RANDOM	3	3(0)	-	-	55.44	52.08(1.948)	-	-	56.1	53.1(1.038)	-	-	54.78	51.01(2.577)	-	-		
	PSO-DS	**3**	3(0)			55.44	51.92(2.85)			53.46	52.82(0.328)			54.78	52(2.652)				
Ecoli(7)	PSO-RANDOM	3	3(0)	-	-	79.46	76.99(2.43)	-	-	80.95	76.11(3.52)	-	-	80.95	76.31(3.93)	-	-	3 [52]	77.38 [52]
	IDS-RANDOM	3	3(0)	-	-	79.46	77.03(2.49)	-	-	80.95	77.4(3.67)	-	-	80.95	77.3(3.7)	-	-		
	PSO-DS	**3**	3(0)			79.46	76.47(3.25)			**80.95**	75.32(4.44)			80.95	75.59(4.55)				
Glass(9)	PSO-RANDOM	2	2(0)	-	-	66.36	58.43(6.18)	-	-	64.95	57.98(6.8)	-	-	68.22	57.9(7.01)	-	-	8.44 [30]	65.14 [30]
	IDS-RANDOM	2	2(0)	-	-	66.36	58.29(7.35)	-	-	63.08	55.41(6.51)	-	-	68.22	55.41(8.07)	-	-		
	PSO-DS	**2**	2(0)			**66.36**	56.81(7.07)			64.95	55.96(8.23)			68.22	57.43(8.7)				
Ionosphere(34)	PSO-RANDOM	2	4.33(1.49)	v	v	88.89	85.35(2.58)	-	-	87.75	86.01(1.69)	-	-	87.75	85.11(2.84)	*	*	7.3 [30]	86.17 [30]
	IDS-RANDOM	4	6.5(1.08)	v	v	88.89	87.27(1.42)	*	*	**91.45**	87.31(2.15)	-	-	89.17	86.82(2.3)	*	*		
	PSO-DS	**2**	**2.67(0.49)**			89.74	83.9(2.89)			89.17	85.47(2.68)			88.89	82.69(2.87)				
Lung(56)	PSO-RANDOM	4	9.83(2.94)	v	v	87.5	72.65(9.62)	-	-	87.5	70.31(11.88)	-	-	87.5	73.17(8.58)	-	-	NA	NA
	IDS-RANDOM	13	14.58(0.9)	v	v	84.37	65.62(8.104)	v	v	87.5	66.15(8.409)	v	v	75	64.32(6.314)	v	v		
	PSO-DS	**3**	**5(1.04)**			**87.5**	77.08(8.25)			87.5	76.03(8.671)			84.37	75.51(8.922)				
Soybean small(35)	PSO-RANDOM	2	2.16(0.39)	-	-	100	99.82(0.61)	-	-	100	99.64(1.22)	-	-	100	99.82(0.61)	-	-	2 [29]	100 [29]
	IDS-RANDOM	3	3.66(0.49)	v	v	100	98.93(2.12)	-	-	100	98.75(2.3)	-	-	100	98.58(2.91)	-	-		
	PSO-DS	**2**	**2.16(0.39)**			**100**	99.29(2.13)			100	99.29(2.93)			100	99.29(2.13)				
Wine(13)	PSO-RANDOM	2	2(0)	-	-	88.2	78.41(8.62)	-	-	88.52	77.27(8.74)	-	-	88.21	78.46(7.96)	-	-	2 [52]	90.44 [52]
	IDS-RANDOM	2	2(0)	-	-	94.94	74.43(10.11)	v	v	90.45	72.56(8.73)	v	v	93.26	73.64(10.16)	v	v		
	PSO-DS	**2**	**2(0)**			**94.94**	81.46(11.1)			90.45	80.52(10.09)			93.82	81.27(10.77)				
LSVT(310)	PSO-RANDOM	113	121(4.53)	v	v	80.16	74.73(3.73)	-	-	83.33	75.59(4.11)	-	-	81.75	75.13(3.52)	-	-	NA	NA
	IDS-RANDOM	119	122.83(1.94)	v	v	79.37	74.66(3.44)	-	-	**86.51**	79.3(3.94)	-	-	81.75	75.46(3.67)	-	-		
	PSO-DS	**14**	**17.75(1.86)**			79.37	75.26(3.66)			84.92	75.6(4.3)			80.16	75.4(3.68)				

Table 4.4: IDS-DS with RST measure: Comparison of reduct size and classification accuracy with statistical t-test, and Wilcoxon test. [Classification accuracy is denoted as 'CA'. Statistical t-test is denoted as 'S'. Wilcoxon test is denoted as 'WT'. Mean Subset Size is denoted as 'MSS'.]

Dataset (Total features)	Feature Selection Method	Feature Subset Size				Classification Accuracy (CA) Classifier : J48				Classifier : JRip				Classifier : PART				Best result reported in literature	
		Best	Mean(s.d.)	S	WT	Best	Mean(s.d.)	S	WT	Best	Mean(s.d.)	S	WT	Best	Mean(s.d.)	S	WT	MSS	CA
Cleveland(13)	PSO-RANDOM	3	3(0)	-	-	55.44	51.89(3.067)	-	-	56.1	53.18(1.032)	-	-	54.78	50.84(2.976)	-	-	7.81 [30]	52.6 [30]
	IDS-RANDOM	3	3(0)	-	-	55.44	52.08(1.948)	-	-	56.1	53.1(1.038)	-	-	54.78	51.01(2.577)	-	-		
	IDS-DS	**3**	3(0)			55.44	52.69(1.581)			**56.1**	53.62(1.538)			54.78	51.39(2.092)				
Ecoli(7)	PSO-RANDOM	3	3(0)	-	-	79.46	76.99(2.43)	-	-	80.95	76.11(3.52)	-	-	80.95	76.31(3.93)	-	-	3 [52]	77.38 [52]
	IDS-RANDOM	3	3(0)	-	-	79.46	77.03(2.49)	-	-	80.95	77.4(3.67)	-	-	80.95	77.3(3.7)	-	-		
	IDS-DS	**3**	3(0)			79.46	77.13(2.46)			**80.95**	76.24(3.78)			80.95	76.34(3.52)				
Glass(9)	PSO-RANDOM	2	2(0)	-	-	66.36	58.43(6.18)	-	-	64.95	57.98(6.8)	-	-	68.22	57.9(7.01)	-	-	8.44 [30]	65.14 [30]
	IDS-RANDOM	2	2(0)	-	-	66.36	58.29(7.35)	-	-	63.08	55.41(6.51)	-	-	68.22	55.41(8.07)	-	-		
	IDS-DS	**2**	2(0)			66.36	60.51(5.31)			63.08	57.91(5.36)			**68.22**	59.58(6.07)				
Ionosphere(34)	PSO-RANDOM	2	4.33(1.49)	v	v	88.89	85.35(2.58)	-	-	87.75	86.01(1.69)	-	-	87.75	85.11(2.84)	-	-	7.3 [30]	86.17 [30]
	IDS-RANDOM	4	6.5(1.08)	v	v	88.89	87.27(1.42)	-	-	**91.45**	87.31(2.15)	-	-	89.17	86.82(2.3)	-	-		
	IDS-DS	**2**	**3(0.43)**			88.89	84.83(2.62)			88.89	84.71(2.3)			90.31	84.05(3.18)				
Lung(56)	PSO-RANDOM	4	9.83(2.94)	v	v	87.5	72.65(9.62)	-	-	87.5	70.31(11.88)	-	-	87.5	73.17(8.58)	-	-	NA	NA
	IDS-RANDOM	13	14.58(0.9)	v	v	84.37	65.62(8.104)	-	-	87.5	66.15(8.409)	-	-	75	64.32(6.314)	-	-		
	IDS-DS	5	**5.92(0.67)**			**87.5**	69(8.983)			87.5	69.78(8.463)			84.37	69.26(9.685)				
Soybean small(35)	PSO-RANDOM	**2**	**2.16(0.39)**	*	*	100	99.82(0.61)	-	-	100	99.64(1.22)	-	-	100	99.82(0.61)	-	-	2 [29]	100 [29]
	IDS-RANDOM	3	3.66(0.49)	v	v	100	98.93(2.12)	-	-	100	98.75(2.3)	-	-	100	98.58(2.91)	-	-		
	IDS-DS	**2**	2.75(0.87)			**100**	98.94(2.13)			100	98.22(3.25)			100	99.11(2.12)				
Wine(13)	PSO-RANDOM	2	2(0)	-	-	88.2	78.41(8.62)	-	-	88.52	77.27(8.74)	-	-	88.21	78.46(7.96)	-	-	2 [52]	90.44 [52]
	IDS-RANDOM	2	2(0)	-	-	94.94	74.43(10.11)	-	-	90.45	72.56(8.73)	-	-	93.26	73.64(10.16)	-	-		
	IDS-DS	**2**	2(0)			**94.94**	76.69(8.32)			90.45	75.23(7.12)			93.26	76.78(8.39)				
LSVT(310)	PSO-RANDOM	113	121(4.53)	v	v	80.16	74.73(3.73)	-	-	83.33	75.59(4.11)	-	-	81.75	75.13(3.52)	-	-	NA	NA
	IDS-RANDOM	119	122.83(1.94)	v	v	79.37	74.66(3.44)	-	-	**86.51**	79.3(3.94)	-	-	81.75	75.46(3.67)	-	-		
	IDS-DS	**18**	**20.67(2.1)**			84.13	75.4(4.69)			80.95	75.13(3.83)			79.37	73.74(4.59)				

Table 4.4 show results of t-test and Wilcoxon test for IDS-DS, it is observed from Table 4.4 that performance of IDS-DS is also comparable to random version of PSO or IDS in terms of statistical significance in case of classification accuracy. The significance of IDS-DS is visible in the reasonably large dataset like *Soybean small*, *Lung* and *LSVT*. In smaller datasets all the methods provide the stable sized reducts, i.e. s.d = 0.

Tables 4.3 and 4.4 show best results reported in literature in terms of Mean Subset size (denoted as MSS) and classification accuracy (denoted as CA). Further, It is evident from Table 4.3 and 4.4 that the performance of PSO-DS and IDS-DS is better than that of state-of-the-art methods suggested in the literature. While using rough dependency measure as fitness function, these method provides smaller reduct with comparable or more accuracy than the existing best method reported in the literature, for all the dataset.

Table 4.5 depicts the statistical analysis of convergence of PSO-DS. It is observed from the Table 4.5 that for all the datasets used in this work, DS initialization based PSO converges in early generation with respect to randomly initialized PSO. Table 4.6 depicts the statistical analysis of convergence of IDS-DS. It is observed from the Table 4.6 that for all the datasets used in this work, DS initialization based IDS converges in early generation with respect to randomly initialized IDS.

It is also observed from the Table 4.6 that, in general, DS initialization based IDS converges in early generations compared to randomly initialized IDS. Especially, in the case of *Ionosphere*, Lung, *Soybean small* and *LSVT* IDS DS converges very fast. Thus, it is observed that DS version of PSO and IDS are superior to their random version. Table 4.7 ranks all the methods discussed above using Friedman ranking. It shows that the DS-initialized version are better than their corresponding random version and PSO-DS is better than IDS-DS.

4.4 PSO and IDS with L-FRFS measure

In this work PSO and IDS have been applied for simultaneous feature selection. The fitness functions have been computed using fuzzy rough dependency measure of L-FRFS suggested in [8,30,32]. All the particles are initialized randomly, and fuzzy lower approximation for all classes are computed using Lukasiewicz fuzzy implicator. In the next step, the fuzzy positive regions, are computed. Consequently fuzzy rough dependency measure

Table 4.5: Statistical analysis of number of generations taken by the dataset for converging to a solution for PSO with RST measure

	PSO-RANDOM			PSO-DS		
Dataset	Min	Mean	s.d.	Min	Mean	s.d.
Cleveland	1	12	14.83	1	3.25	2.59
Ecoli	1	8.33	10.81	1	1.25	0.45
Glass	1	13.66	11.72	1	1.08	0.28
Ionosphere	2	19.41	10.17	1	1.58	1.44
Lung	1	30.5	17.23	1	3.41	1.78
Soybean Small	7	20.33	11.64	1	5.91	5.86
Wine	1	1.58	1.16	1	1	0
LSVT	37	79.16	19.32	11	16.33	13.5

Table 4.6: Statistical analysis of number of generations taken by the dataset for converging to a solution for IDS with RST measure

	IDS-RANDOM			IDS-DS		
Dataset	Min	Mean	s.d.	Min	Mean	s.d.
Cleveland	1	8.58	10.63	1	8.16	10.89
Ecoli	1	1	0	1	1.08	0.28
Glass	1	1	0	1	1.16	0.38
Ionosphere	3	33.33	21.33	1	1	0
Lung	10	40.16	19.33	1	1	0
Soybean Small	1	26.83	17.23	1	9.16	15.4
Wine	1	2.41	1.67	1	1	0
LSVT	3	48.67	34.92	1	1	0

Table 4.7: Friedman ranking with RST measure

Methods	FR	Rank
PSO-DS	2.2969	1
PSO-RANDOM	2.4219	2
IDS-DS	2.5000	3
IDS-RANDOM	2.7813	4

is computed and used as a part of the fitness function.

4.4.1 Experiments, Results and Discussions

Parameters and assumptions considered regarding experiments are discussed as follows.

Data Normalization

In this work the data discussed in Section 4.3.1 has been scaled in the range [0,1], using the following equation.

$$E_i^{norm} = \frac{E_i^t - E_i^{min}}{E_i^{max} - E_i^{min}} \tag{4.3}$$

where E_i^{norm} is a normalized value of the i^{th} element of a given feature, E_i^{min} and E_i^{max} are respectively minimum and maximum values of all the elements of the corresponding feature.

Further, Fuzzy Similarity Matrices (FSMs) have been computed separately using Equation (3.19) for each of the individual feature of the corresponding dataset individually. These FSMs are utilized to compute dependency measure corresponding to the feature set and combination of feature sets.

Parameter Setting

Parameters used for PSO [4] referenced in Equation (1.1) are as follows. C_1 and C_2 will have their frequently used value 2. The inertia weight, w decreases from 1.4 to 0.4 and

$$w = (w - 0.4) * \left(\frac{Max\_iter - Current\_iter}{Max\_iter} \right) + 0.4 \tag{4.4}$$

Parameters used for IDS [29], are set as; $C_w = 0.1$, $C_p = 0.4$, $C_g = 0.9$.

Fitness value

The value of fitness will depend upon fuzzy rough dependency measure $\gamma'_P(Q)$ as shown in Equation (3.21). The value of fitness is computed using the following formula [4].

$$Fitness\_Value = 0.9 * \gamma'_P(Q) + 0.1 * \left(\frac{|C| - |R|}{|C|}\right) \tag{4.5}$$

where $\gamma'_P(Q)$ is the dependency measure of the selected reduct, R. C and D are respectively total number of features and decision feature in the dataset.

Classification Accuracy

In this work J48 [78], JRip [79] and PART [80] classifiers are used on each of the dataset, to calculate S-10FCV classification accuracy for comparison purposes.

Results and Discussions

PSO-RANDOM, IDS-RANDOM, PSO-DS and IDS-DS have been executed 12 times for each of the benchmark datasets. Each run is of 100 generations with a population size of 100. Classification accuracies are computed using J48, JRip and PART classifiers in terms of their best, mean and s.d. values. Statistical t-test (denoted 'S'), Wilcoxon test ('WT') and Friedman ranking are also performed for classification accuracy and for the reduct size, with respect to PSO-DS and IDS-DS. Table 4.8 shows results of t-test, and Wilcoxon test and Friedman ranking for PSO-DS method and Table 4.9 shows results of t-test and Wilcoxon test for IDS-DS in terms of classification accuracy. In these tables, statistical significance of PSO-DS and IDS-DS are compared to the PSO-random and the IDS-random has been tabulated. From these values, it is evident that the proposed PSO-DS and IDS-DS are always comparable or better in terms of statistical significance for all the three classifiers used in this work.

It is observed from Tables 4.8 and 4.9, that fuzzy rough set methodology facilitates the reduction in the size of the feature subset, with acceptable and comparable classification accuracies. Resulting reducts yield high fuzzy rough dependency measures even with the reduced size features, in the reduct. The significance of DS-initialization is visible in the reasonably large dataset like *Soybean small*, *Lung* and *LSVT*.

Table 4.8: PSO-DS with L-FRFS measure: Comparison of reduct size and classification accuracy with statistical t-test, and Wilcoxon test. [Classification accuracy is denoted as 'CA'. Statistical t-test is denoted as 'S'. Wilcoxon test is denoted as 'WT'. Mean Subset Size is denoted as 'MSS'.]

Dataset (Total features)	Feature Selection Method	Feature Subset Size				Classification Accuracy (CA)												Best result reported in literature	
						Classifier : J48				Classifier : JRip				Classifier : PART					
		Best	Mean(s.d.)	S	WT	Best	Mean(s.d.)	S	WT	Best	Mean(s.d.)	S	WT	Best	Mean(s.d.)	S	WT	MSS	CA
Cleveland(13)	PSO-RANDOM	6	6.26(0.46)	-	v	52.47	52.09(0.7)	-	v	53.79	53.54(0.5)	-	v	52.14	51.24(1.89)	-	v	7.81 [30]	52.6 [30]
	IDS-RANDOM	6	6.51(0.53)	v	v	52.47	51.69(0.8)	-	v	53.79	53.52(0.29)	-	v	52.14	51.14(1.87)	-	v		
	PSO-DS	**6**	**6(0)**			52.47	52.47(0)			**53.79**	53.79(0)			52.14	52.14(0)				
Ecoli(7)	PSO-RANDOM	5	5(0)	-	-	82.44	82.44(0)	-	-	81.25	81.25(0)	-	-	80.65	80.65(0)	-	-	3 [52]	77.38 [52]
	IDS-RANDOM	5	5(0)	-	-	82.44	82.44(0)	-	-	81.25	81.25(0)	-	-	80.65	80.65(0)	-	-		
	PSO-DS	**5**	5(0)			**82.44**	82.44(0)			81.25	81.25(0)			80.65	80.65(0)		-		
Glass(9)	PSO-RANDOM	8	8(0)	-	-	64.49	64.49(0)	-	-	69.16	69.16(0)	-	-	68.69	68.69(0)	-	-	8.44 [30]	65.14 [30]
	IDS-RANDOM	8	8(0)	-	-	64.49	64.49(0)	-	-	69.16	69.16(0)	-	-	68.69	68.69(0)	-	-		
	PSO-DS	**8**	8(0)			64.49	64.49(0)			**69.16**	69.16(0)			68.69	68.69(0)				
Ionosphere(34)	PSO-RANDOM	6	7.34(0.89)	-	v	91.74	89.51(1.34)	-	v	91.45	89.63(1.24)	-	-	91.46	89.58(1.67)	-	-	7.3 [30]	86.17 [30]
	IDS-RANDOM	7	8.17(0.72)	-	v	**93.73**	90.2(2.05)	-	*	92.02	89.18(1.87)	-	-	93.73	89.37(2.42)	-	v		
	PSO-DS	**6**	**7(0.6)**			91.74	89.57(1.72)			91.74	89.06(1.78)			91.7	89.53(1.53)				
Lung(56)	PSO-RANDOM	5	6.51(1.32)	v	v	71.87	65.89(3.88)	v	v	78.12	66.93(4.89)	v	v	84.37	66.91(8.26)	v	v	NA	NA
	IDS-RANDOM	13	13.42(0.68)	v	v	87.5	71.62(8.99)	-	v	84.37	72.14(8.66)	-	v	81.25	66.13(12.83)	-	v		
	PSO-DS	**4**	**4.91(0.79)**			**87.5**	74.2(7.77)			87.5	76.28(7.1)			87.5	75.26(7.94)				
Soybean small(35)	PSO-RANDOM	2	2.59(0.5)	-	v	100	99.47(1.33)	-	v	100	99.12(1.43)	-	v	100	99.63(0.81)	-	-	2 [29]	100 [29]
	IDS-RANDOM	3	3.75(0.63)	v	v	100	99.12(1.92)	-	v	100	98.94(1.44)	-	v	100	99.28(1.87)	-	v		
	PSO-DS	**2**	**2.25(0.45)**			**100**	99.65(0.82)			100	99.65(0.83)			100	99.65(0.83)				
Wine(13)	PSO-RANDOM	4	4(0)	-	-	93.82	91.55(2.93)	-	v	92.13	90.39(1.7)	-	v	93.82	91.74(2.42)	-	v	2 [52]	90.44 [52]
	IDS-RANDOM	4	4(0)	-	-	93.82	93.15(0.67)	-	-	92.13	90.62(1.03)	-	v	93.82	92.73(1.22)	-	-		
	PSO-DS	**4**	4(0)			**93.82**	93.08(0.81)			92.13	90.92(1.07)			93.82	92.77(1.13)				
LSVT(310)	PSO-RANDOM	113	120.95(4.35)	v	v	80.15	75.13(3.31)	-	*	82.53	77.16(2.67)	-	*	81.75	75.39(3.84)	-	*	NA	NA
	IDS-RANDOM	119	122.8(1.98)	v	v	79.37	74.96(3.13)	-	*	**86.51**	78.72(2.96)	-	*	81.75	75.21(3.33)	-	*		
	PSO-DS	**13**	**16.51(2.12)**			82.54	73.15(5.13)			80.95	74.21(3.94)			80.95	73.42(5.52)				

Table 4.9: IDS-DS with L-FRFS measure: Comparison of reduct size and classification accuracy with statistical t-test, and Wilcoxon test. [Classification accuracy is denoted as 'CA'. Statistical t-test is denoted as 'S'. Wilcoxon test is denoted as 'WT'. Mean Subset Size is denoted as 'MSS'.]

Dataset (Total features)	Feature Selection Method	Feature Subset Size				Classification Accuracy (CA)												Best result reported in literature	
						Classifier : J48				Classifier : JRip				Classifier : PART					
		Best	Mean(s.d.)	S	WT	Best	Mean(s.d.)	S	WT	Best	Mean(s.d.)	S	WT	Best	Mean(s.d.)	S	WT	MSS	CA
Cleveland(13)	PSO-RANDOM	**6**	**6.26(0.46)**	-	*	52.47	52.09(0.7)	-	*	53.79	53.54(0.5)	-	-	52.14	51.24(1.89)	-	*	7.81 [30]	52.6 [30]
	IDS-RANDOM	6	6.51(0.53)	v	v	52.47	51.69(0.8)	-	-	53.79	53.52(0.29)	-	-	52.14	51.14(1.87)	-	*		
	IDS-DS	**6**	6.41(0.51)			52.47	51.8(0.82)			**53.79**	53.53(0.3)			52.14	50.41(2.39)				
Ecoli(7)	PSO-RANDOM	5	5(0)	-	-	82.44	82.44(0)	-	-	81.25	81.25(0)	-	-	80.65	80.65(0)	-	-	3 [52]	77.38 [52]
	IDS-RANDOM	5	5(0)	-	-	82.44	82.44(0)	-	-	81.25	81.25(0)	-	-	80.65	80.65(0)	-	-		
	IDS-DS	**5**	5(0)			**82.44**	82.44(0)			81.25	81.25(0)			80.65	80.65(0)				
Glass(9)	PSO-RANDOM	8	8(0)	-	-	64.49	64.49(0)	-	-	69.16	69.16(0)	-	-	68.69	68.69(0)	-	-	8.44 [30]	65.14 [30]
	IDS-RANDOM	8	8(0)	-	-	64.49	64.49(0)	-	-	69.16	69.16(0)	-	-	68.69	68.69(0)	-	-		
	IDS-DS	**8**	8(0)			64.49	64.49(0)			**69.16**	69.16(0)			68.69	68.69(0)				
Ionosphere(34)	PSO-RANDOM	**6**	**7.34(0.89)**	-	*	91.74	89.51(1.34)	-	v	91.45	89.63(1.24)	-	-	91.46	89.58(1.67)	-	*	7.3 [30]	86.17 [30]
	IDS-RANDOM	7	8.17(0.72)	-	v	93.73	90.2(2.05)	-	*	92.02	89.18(1.87)	-	-	93.73	89.37(2.42)	-	*		
	IDS-DS	7	7.83(0.57)			**93.73**	89.87(2.16)			92.02	89.32(1.83)			93.73	89.06(2.4)				
Lung(56)	PSO-RANDOM	5	6.51(1.32)	v	v	71.87	65.89(3.88)	v	v	78.12	66.93(4.89)	v	v	84.37	66.91(8.26)	v	v	NA	NA
	IDS-RANDOM	13	13.42(0.68)	v	v	87.5	71.62(8.99)	-	v	84.37	72.14(8.66)	-	v	81.25	66.13(12.83)	-	v		
	IDS-DS	**5**	**6.25(0.62)**			**87.5**	78.13(7.18)			87.5	75.25(10.94)			87.5	74.74(8.05)				
Soybean small(35)	PSO-RANDOM	**2**	**2.59(0.5)**	-	*	100	99.47(1.33)	-	-	100	99.12(1.43)	-	v	100	99.63(0.81)	-	-	2 [29]	100 [29]
	IDS-RANDOM	3	3.75(0.63)	v	v	100	99.12(1.92)	-	v	100	98.94(1.44)	-	v	100	99.28(1.87)	-	v		
	IDS-DS	**2**	2.75(0.62)			**100**	99.29(1.38)			100	99.64(0.82)			100	99.64(0.82)				
Wine(13)	PSO-RANDOM	4	4(0)	-	-	93.82	91.55(2.93)	-	v	92.13	90.39(1.7)	-	-	93.82	91.74(2.42)	-	v	2 [52]	90.44 [52]
	IDS-RANDOM	4	4(0)	-	-	93.82	93.15(0.67)	-	v	92.13	90.62(1.03)	-	*	93.82	92.73(1.22)	-	-		
	IDS-DS	**4**	4(0)			**93.82**	93.55(0.51)			91.57	90.33(0.75)			93.82	93.12(1.26)				
LSVT(310)	PSO-RANDOM	113	120.95(4.35)	v	v	80.15	75.13(3.31)	-	v	82.53	77.16(2.67)	-	*	81.75	75.39(3.84)	-	*	NA	NA
	IDS-RANDOM	119	122.8(1.98)	v	v	79.37	74.96(3.13)	-	*	**86.51**	78.72(2.96)	-	*	81.75	75.21(3.33)	-	*		
	IDS-DS	**16**	**20.76(2.87)**			80.95	75.54(4.3)			83.33	75.15(3.84)			78.57	72.61(3.75)				

For example in case of *LSVT*, PSO-DS gives 16.51 reduct size, and IDS-DS gives 20.76 reduct size, as compared to 120.95 reduct size given by PSO-random and 122.8 given by IDS-random, where the unreduced feature size is 310, and the feature subsets(reducts) evaluated from all these method produce the comparable classification accuracy.

Similar is the case of *Lung* dataset, where PSO-DS and IDS-DS produce the smaller reducts with high classification accuracies, when compared to random version (Tables 4.8 and 4.9). In *Soybean small*, accuracies achieved are very near to 100 percent with smallest reducts. In datasets *Cleveland* and *Ionosphere* also, PSO-DS and IDS-DS reduce the size of the reduct with acceptable classification accuracies. In smaller datasets *Ecoli*, *Glass* and *Wine*, all the methods provide the stable sized reducts, i.e. s.d = 0.

Through these investigations, it has been demonstrated that in almost all of the cases, classification accuracy is acceptable. It is observed from Tables 4.8 and 4.9 that PSO-DS and IDS-DS are also always comparable to or better than PSO-RANDOM and IDS-RANDOM in terms of statistical significance of classification accuracy.

Tables 4.8 and 4.9 show best results reported in literature in terms of Mean Subset size (denoted as MSS) and classification accuracy (denoted as CA). Further, It is evident from the Tables 4.8 and 4.9 that the performance of PSO-DS and IDS-DS is better than that of state-of-the-art methods suggested in literature for the dataset *Cleveland*, *Glass* and *Ionosphere*. While using fuzzy rough dependency measure as fitness function, these method provides better accuracy for *Ecoli* and *Wine* dataset at the cost of relatively large reduct size than the existing best method reported in literature.

Thus, it is established that optimal reducts obtained using PSO-DS and IDS-DS, having maximum possible fuzzy rough dependency measure have acceptable accuracies, with relatively smaller reduct size than in the case of PSO-RANDOM and IDS-RANDOM.

The following qualitative reasons explains the smaller feature subset of RST method compared to L-FRFS approach.

1. L-FRFS is not used for improved performance in terms of accuracy or reduct size, rather it was used for capturing fuzziness in description of datasets. Hence, in the effort to capture fuzziness, reduct may not be smaller as compared to a rough set approach.

Fuzzy lower approximation maintains dependency of data which will never be zero, whereas, rough set is crisp and hard-limits the approximation. Due to this, RST may ignore or preserve features in a crisp manner resulting in loss of dependency and therefore

Table 4.10: Statistical analysis of number of generations taken by the dataset for converging to a solution for PSO with L-FRFS measure

	PSO-RANDOM			PSO-DS		
Dataset	Min	Mean	s.d.	Min	Mean	s.d.
Cleveland	2	13.67	25.43	3	8.16	5.5
Ecoli	1	1.58	0.67	1	1.75	0.62
Glass	1	1.91	0.67	1	1.08	0.28
Ionosphere	1	14.16	12.91	2	8.16	10.72
Lung	24	45.16	9.87	1	2.58	1.67
Soybean Small	8	12.67	5.53	1	3.33	2.34
Wine	1	9.67	16.87	1	16.75	14.83
LSVT	72	84.41	8.24	1	14.83	8.5

reduct produced may be smaller in size.

2. In RST, there is a need to discretize the feature values. The discernibility of features are affected by the quantization and therefore become dependent on the quantization of feature value. Whereas, in L-FRFS the real feature values are taken as it is and no quantization is required and discernibility of features are therefore more accurate as well as meaningful. This is the reason why the RST approach may provide smaller size features compared to L-FRFS approach.

It is observed from the Table 4.10 that for all the dataset used in this work, PSO-DS converges in early generation compared to PSO-RANDOM. It is also observed from the Table 4.11 that in general, IDS-DS converges in early generation compared to IDS-RANDOM. It is observed that especially in the case of *Soybean small* and *LSVT*, IDS-DS converges very fast.

Table 4.12 shows ranking of the above discussed methods obtained from Friedman ranking. It is observed from these ranking that performance of PSO-DS and IDS-DS are superior with respect to PSO-RANDOM and IDS-RANDOM.

Table 4.11: Statistical analysis of number of generations taken by the dataset for converging to a solution for IDS with L-FRFS measure

	IDS-RANDOM			IDS-DS		
Dataset	Min	Mean	s.d.	Min	Mean	s.d.
Cleveland	2	44.41	30.3	2	48.91	33.14
Ecoli	1	2.08	1.5	1	2.41	2.06
Glass	4	11.58	8.9	1	3.41	6.06
Ionosphere	8	35.41	22.15	1	41.91	32.09
Lung	4	37.75	20.32	1	1	0
Soybean Small	5	45.33	34.37	1	2.67	5.77
Wine	14	53.75	25.07	10	42.75	28.46
LSVT	1	48.16	35.22	1	1	0

Table 4.12: Friedman ranking with L-FRFS measure

Methods	FR	Rank
PSO-DS	2.250	1
IDS-DS	2.750	2
PSO-RANDOM	2.969	3
IDS-RANDOM	3.281	4

4.5 Conclusion

Due to distributed sampled seed population, DS initialized swarm algorithms always produce smaller reducts, as they are able to pick the appropriate reducts in early generation. Following conclusion can be drawn.

1. Using RST measure, PSO-DS and IDS-DS, in general, achieve better performance as compared to PSO-RANDOM and IDS-RANDOM. PSO-DS and IDS-DS, outperform PSO-RANDOM and IDS-RANDOM in such cases where feature reduction is actually sought after, i.e. datasets having large number of features. PSO-DS and IDS-DS, outperform PSO-RANDOM and IDS-RANDOM for large datasets, without compromising with classification accuracy.

2. Using L-FRFS measure, PSO-DS and IDS-DS, in general, achieve better performance as compared to PSO-RANDOM and IDS-RANDOM, which has been established using t-test, Wilcoxon test and Friedman test.

3. It is observed from Friedman test, that using RST measure, rank of PSO-RANDOM is better than that of IDS-DS, but with L-FRFS measure rank of IDS-DS is better than that of PSO-RANDOM. Other methods namely IDS-RANDOM and PSO-DS preserve their ranking order of performance, regardless of feature selection measures used.