AN OPTIMIZED CLASSIFICATION USING COMBINATION OF IMPROVED
ADAPTIVE FUZZY PARTICLE SWARM OPTIMIZATION ALGORITHM WITH
THE ROUGH THEORY

Hamid Tabatabaee\textsuperscript{a},* , Parisa Molavi\textsuperscript{b}, Maryam Baratzadeh\textsuperscript{b}, NasibehRadyRaz\textsuperscript{c}

\textsuperscript{a} Young Researchers and Elite Club, Quchan Branch, Islamic Azad University, Quchan, Iran.
\textsuperscript{b} Department of Computer engineering, Mashhad branch, Islamic Azad University, Mashhad, Iran.
\textsuperscript{c} Young Researchers and Elite Club, Mashhad Branch, Islamic Azad University, Mashhad, Iran.

ABSTRACT: Classification is one of the widely issues used in data mining. Today, due to dealing with large data sets, finding a good way to promote classification's accuracy is considered as major issues. Using all the available properties in the data sets not only increase the complexity but also reduce the accuracy of the classification. The use of evolutionary algorithms instead of the complete search methods is a solution. Due to the fast convergence of the PSO algorithm, the chance of falling into local optimum is increased. The proposed algorithm is able to improve the efficiency of the search process, with dynamic parameters based on fuzzy logic. FPSO algorithm approach is a way to make adaptive PSO parameters dynamically instead of holding constant weight. In this paper, first a new technique is defined by making the parameter weight dynamically based on the fuzzy logic in order to increase diversity and to establish tradeoff between local exploration and global exploration. Then, in order to have an optimized classification, the combination of this method with Rough set for feature selection is introduced. The first approach is used for optimizing a few number of benchmark mathematical functions and the second approach is used for finding optimized classification in some of the UCI data sets. Results show the merits of the proposed method in comparison to the FPSO method and the three other classification methods.

Keywords- Classification; Particle Swarm Optimization, Fuzzy Logic, Rough Theory.

Symbols and variables Tables

<table>
<thead>
<tr>
<th>Symbol/variable</th>
<th>meaning</th>
<th>Symbol/variable</th>
<th>meaning</th>
<th>Symbol/variable</th>
<th>meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>$U$</td>
<td>Universal Set</td>
<td>$BND_p$</td>
<td>Boundary Region</td>
<td>$W$</td>
<td>Weight</td>
</tr>
<tr>
<td>$X$</td>
<td>Subset of U</td>
<td>$PX$</td>
<td>Low approximation from X</td>
<td>$DV_{avg}$</td>
<td>Average proportion of time</td>
</tr>
<tr>
<td>$A$</td>
<td>Non-empty finite set of attributes (features)</td>
<td>$\bar{PX}$</td>
<td>High approximation from X</td>
<td>$V_{max}$</td>
<td>Maximum speed of particle</td>
</tr>
<tr>
<td>$a$</td>
<td>Member of $A$</td>
<td>$C$</td>
<td>Conditional attribute set</td>
<td>$P_i$</td>
<td>Particle with I index</td>
</tr>
<tr>
<td>$I$</td>
<td>Information System</td>
<td>$D$</td>
<td>Attribute decision set</td>
<td>$P_{gbest}$</td>
<td>Best Particle</td>
</tr>
<tr>
<td>$x, y$</td>
<td>Data variables and member of $U$</td>
<td>$\gamma_{(D)}$</td>
<td>Quality of classification</td>
<td>$\alpha$</td>
<td>Weight of Rough set</td>
</tr>
<tr>
<td>$P, Q$</td>
<td>Equivalence relations</td>
<td>$Re d$</td>
<td>Set of all reduced feature</td>
<td>$\beta$</td>
<td>Weighted value for the minimum length of attributes</td>
</tr>
<tr>
<td>$F_a$</td>
<td>Mapping Function</td>
<td>$Core$</td>
<td>Share of all reduced feature</td>
<td>$B$</td>
<td>Selected features</td>
</tr>
<tr>
<td>$V_a$</td>
<td>A set of values of a Positive region</td>
<td>$IND$</td>
<td>Division of the World Set</td>
<td>$C$</td>
<td>Number of Features</td>
</tr>
<tr>
<td>$POS_p$</td>
<td>Positive region</td>
<td>$C_1$</td>
<td>Personal learning coefficient</td>
<td>$K_{\text{max}}$</td>
<td>Maximum number of iterations</td>
</tr>
<tr>
<td>$NEG_p$</td>
<td>Negative Region</td>
<td>$C_2$</td>
<td>Social learning coefficient</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

May-June
1. INTRODUCTION
Dealing with the large data sets, a good way to improve the accuracy of classification is one of the most important issues needed to be resolved. Using all the features in the data set may increase the computation time and can reduce the accuracy of classification. Thus, feature selection is one of the issues that found applications in machine learning and signal processing.

Feature selection is the selection of the appropriate feature subset on the basis of the main features of a dataset. The following selection must be desirable to the extent that the resulting classification accuracy compared to use all the characteristics of the data set the same or even desired. Select the appropriate subset, in dealing with datasets with incomplete information, noisy or ambiguous frequently seen in real-world problems is desperately needed.

Rough set theory [1-3] has the ability to deal with uncertain and vague data. This theory can be used to select desired features and also able to predict the implications of the decision as well as the basic feature set. In other words, the application of this theory is to select the feature that can find the minimum number of features, but with high classification accuracy [4]. The number of rules generated by the property is considered as a criterion [1].

Rough theory is one of the most powerful methods to select features but it is still unable to find optimal subset. For this purpose, optimization techniques would be used in order to find the optimal subset [5]. For classification and feature selection by the theory of Rough, many methods have been proposed. In the following brief overview several methods have been carried out.

2. LITERATURE BACKGROUND
Starzyk et al. [6] presented a method based on the strong equivalence in order to simplify the indistinguishable functions. In this method, a costly solution to the problem is found that only works for simple data collection. It should be mentioned that size Reduction problem is considered as a NP-Hard problem. Since considering all the features of the problem lead to an exponential type, thus heuristic approaches are examined [7].

Hill-climbing methods and evolutionary optimization algorithms can be considered as the example of the greedy techniques and random methods respectively. In the hill-climbing method, feature selection is done, based on the forward selection or backward elimination.

Hu et al. [8] used reduction algorithm that is based on the feature selection with positive region constrain (features that do not cause contradiction in the decision feature) that guide the heuristic algorithm. In order to reduce the features, both distinguishable and indistinguishable relationships are considered by Susmaga et al. in [9, 10]. Hill climbing, which is considered as heuristic method is not guarantee finding the lowest or best reduced features. Hence, there is no heuristic to guarantee the complete optimality.

Recently, population-based evolutionary method is used for Rough set feature selection [11]. Wang [12] used genetic algorithms in order to find the smallest subset. In this method, a genetic algorithm is combined with a greedy algorithm in order to generate a minimal subset. Time cost and no guarantee in reducing the result are two problems of this method. Zhai et al. [13] presented an integrated feature selection based on the Rough theory and genetic algorithm. Skowron et al. in [14, 15] used Rough set and genetic algorithm for feature selection. In their method, after calculating the upper and lower approximation, the training data can be divided into definitive training data and possible training data, then the evolutionary algorithms will be covered the best rules from the data collection and the fitness function is considered as the quality of the extracted rules. Jensen et al. in [16] has found that the minimum reduction of rough set based on ant colony optimization algorithm (ACO).

In the [17], Xia et al. used frog algorithm for feature selection based on the entropy weighting method. In [18], a new hybrid genetic algorithm based on the local search of the Rough theory is presented. In this method, all feature subsets including key features that have been developed in an evolutionary process, accelerated the convergence. In [19] a genetic programming based on the filter method is presented to be a multi objective method for feature selection in binary classification problems.

In [20] a feature selection method based on the ant colony optimization and fuzzy rough theory method is presented for web Content classification problem and complex system monitoring. Unler et al. [21] used a PSO algorithm based on the feature selection that is used an adaptive selection strategy.

Particle Swarm Optimization (PSO) algorithm is an evolutionary method that inspired from the social animals like flogs of birds and school of fish [24, 25]. This algorithm used the flying birds’ method in information exchanging process. Each particle consider as a potential solution with definitive velocity in problem space. Particles will set their positions based on their best personal position and the best global position. Particles which are existed in the best position will be considered as the results. PSO is used for a variety of optimization problems. Researches show the merits of this algorithm in comparison to the genetic algorithm [24, 25]. Thus, we use PSO for feature selection.

In [27] an optimized fuzzy classification method using a PSO with dynamic parameters is presented. The organization of the rest of this paper is as follows: Section three is presented the basic concepts. The proposed approach is presented in section four. In section five simulations are presented. Comparison and results will be presented in section six. Finally the section seven contains the conclusion and the future works.

3. BASIC CONCEPT
3.1 Rough Theory
The rough theory is a mathematical approach to deal with the uncertainty, ambiguity and imprecision [1, 3]. Rough deals with the analysis of data tables and its aim is to obtain an approximation of the acquired data. It eliminates redundant data without loss of essential data collection. As a result of the reduction of data, a set of meaningful rules for easier
decision making will be produced. In fact, it can be stated that the rough set is a mapping of the raw data space to the concepts space by reducing the data space and choosing key features.

Rough set theory has several points in common with the theory of fuzzy sets theory, intuitionist fuzzy, discriminate analysis using Boolean logic. However, rough set theory can be considered as an independent theory [28].

Objects are characterized by the same information from the standpoint of indistinguishable available information. The obtained indiscernible relationship of (causal relationship) is a bases of mathematical Rough sets theory. Each set of indistinguishable objects in a collection called fundamental set, and each group of fundamental set is called crisp set. Otherwise, the set will be called vague and imprecise rough sets. A rough set is approximation of the vague concepts that, can be expressed with the crisp pair of lower and upper approximations. The lower approximation is a set that certainly belong to the subset of interest, while the upper approximation is the set of all objects that may belong to the subset. For a given set of features, a set is a rough set if and only if the lower and upper approximations are not equal to each other. The main advantage of Rough set is that except to the provided data, it does not require any additional knowledge. Rough sets do the feature selection only by using the granularity of data[29].

\[ I = (U, A) \] is an information system in which \( U \) is the universal, non-empty and finite set of features. \( A \) is a non-empty and finite set of features. \( \forall a \in A \) is a function that defined \( f_a: U \rightarrow V_a \) in which \( V_a \) is a set of values of \( a \). If \( P \subseteq A \) then there is an equivalence relation as follows:

\[ \text{IND}(P) = \{(x, y) \in U \times U | \forall a \in P, f_a(x) = f_a(y)\} \]  (1)

\( U \) part which is produced by \( \text{IND}(P) \) means \( U/P \). If \( (x, y) \in \text{IND}(P) \), then \( x \) and \( y \) are indistinguishable with features of \( P \). Equivalence classes of the relation are defined by \( P \) [x]. Detection of indistinguishable relation is a mathematical basis of the Rough Sets.

If \( X \subseteq U \), then the lower approximation is \( (pX)p \) and the upper approximation is \( (\bar{p}X)p \) and they will be defined as follows:

\[ pX = \{x \in U | [x]_P \subseteq X\} \quad (2) \]

\[ \bar{p}X = \{x \in U | [x]_P \cap X \neq \emptyset\} \quad (3) \]

If \( Q \subseteq A \) and \( P \) is the equivalence relation, then the positive, negative and boundary regions will be defined as follows:

\[ \text{POS}_P(Q) = \bigcup_{x \in U/Q} PX \quad (4) \]

\[ \text{NEG}_P(Q) = \bigcup_{x \in U/Q} \bar{PX} \quad (5) \]

\[ \text{BND}_P(Q) = \bigcup_{x \in U/Q} \bar{PX} - \bigcup_{x \in U/Q} PX \quad (6) \]

The positive region from the \( U/P \) section is in relation to set of \( P \) which is shown by \( \text{POS}_P(Q) \) where \( U \) is a set of all the elements and it is defined in an absolute way by parts with the \( P \) set and classified with \( U/P \) parts. A set is rough if boundary region is non-empty. Important outcome of the analysis is the discovering of the dependencies between features (attributes). Dependence is defined in the following way: For \( \subseteq A, P \) set is completely dependent to \( Q \) set if and only if \( \text{IND}(P) \subseteq \text{IND}(Q) \). And it means that the part which is generated by a \( P \) set is less than the part which is generated by \( Q \) set. It is also such that any subset of the \( P \) set with degree of \( k \), where \( 0 \leq k \leq 1 \) is dependent. It is shown as \( P \Rightarrow Q_k \) and equal to:

\[ k = \gamma_P(Q) = \frac{|\text{POS}_P(Q)|}{|U|} \quad (7) \]

If \( k=1 \), then the set of \( Q \) is completely dependent to set of \( P \) and if \( 0 < k < 1 \) then set of \( Q \) is approximately dependent to \( P \) set and if \( k=0 \), then \( Q \) set is not dependent to \( P \) set. In other words, if using \( P \) set makes all the elements of the general set of \( U \) to be classified in to \( U/P \) parts, then \( Q \) set completely dependent to \( P \) set.

In a decision system, set of attributes include the set of attribute of \( C \) position and decision set of \( D \). It means that \( A = C \cup D \). Decry of dependency between position and decision attributes is defined by \( \gamma_C(D) \) which is called a quality attributes of classification [18]. The aim of attribute reduction is to remove redundant features so that a same quality classification of the initial set will be provided. A reduction is defined as a subset of \( R \) from the position attribute of \( C \) where \( \gamma_R(D) = \gamma_C(D) \). A decision table may contain several reduction attributes where a set of reduction attributes defined as[1]:

\[ \text{Red} = \{ R \subseteq C | \gamma_R(D) = \gamma_C(D) \forall B \subseteq R, \gamma_B(D) \neq \gamma_C(D) \} \quad (8) \]

In a rough set, a reduced subset is a search for reduced feature with the lowest cardinality which is defined by:

\[ \text{Red}_{d_{\text{min}}} = \{ R \in \text{Red} | \forall R' \in \text{Red}, |R| \leq |R'| \} \quad (9) \]

Unity of all the reductions called core which include elements that cannot be removed. The core is defined as follows:

\[ \text{Core}(C) = \bigcap \text{Red} \quad (10) \]

Rough theory has several applications in engineering, medical data analysis, image processing, and etc. Some of the advantages of Rough set theory is as follows [28]:

- An effective algorithm for finding hidden patterns in data.
- Finding a minimal set of data (data reduction).
- Evaluation of data.
- Production of a minimal set of decision rules from data.
- Simple, easy to understand and interpret the results of the algorithm.
3.2 Basic Definition of FPSO

The PSO algorithm will be defined by two equation velocity and position that presented in (11) and (12).

\[ v_{id} = w \cdot v_{id} + c_1 \cdot \text{Rand}(\cdot) \cdot (p_{gd} - x_{id}) + c_2 \cdot \text{Rand}(\cdot) \cdot (p_{gb} - x_{id}) \] (11)

\[ x_{id} = x_{id} + v_{id} \] (12)

Where w is the inertia weight, vid is the speed of each particle, c1 is the personal learning parameter, c2 is the social learning parameter, pgd is the global best position, xid is the personal best position and Rand() is a random number defined in [0 1]. These parameters derived from [26].

In [27] a fuzzy logic is used in order to dynamically define C1 and C2 parameters for increasing efficiency.

Three fuzzy inference systems will be defined as below:

- The first system used two criteria of repetition and diversity, the second system used repletion and error and the third system used repetition, diversity and error as inputs of fuzzy inference system and c1, c2 as outputs.

4. THE PROPOSED AFPSO ALGORITHM

The flaws of paper [27] are listed below:

1. Increasing in the number of rules.
2. Increasing complexity due to considering extra features in high dimensional data sets.
3. Complexity in coding membership functions for each feature considering range limitation.
4. There should be a trade of between exploration and exploitation by using W (weight) [30], but omitting this elements, the presented PSO in [27] lost diversity and trapped in to local minima.

Therefore to improve the above flaw, we presented the below ideas:

- First, we introduce a dynamic weight parameter using fuzzy logic and define a proper value for it. Second, combining the first idea with the rough theory set, we do the rules reduction and complexity reduction in coding in order to improve the quality of the classification comparing to [30].

4.1 An Improved AFPSO, A New Proposal for using in benchmark mathematical

Regulating w, c1 and c2 have a great effect on the performance of PSO. In [27] c1 and c2 parameters are defined using a fuzzy logic and the value of w (weight) is considered to be fixed. The weight inertia prepare a trade of between local and global exploration and inappropriate weight value decrease the search diversity and caused it to fall in to local minima. In other word, if the particles situated close to the global minima, will go away from that place and if the low value considered for the weight, and the particles will situated in a more distance from the global minima, exploration process use a shorter steps and the chance of falling in local optimization is become very high [30].

Considering these facts, the proposed approach is used the combination of ideas presented in [27,30]. To do this, equation (13) and (14) which are presented in [30] are used for fuzzifying the weight intertie and c1 and c2 presented in [27] for defining an adaptive fuzzy particle swarm optimization. The proposed method is implemented on the FOSZO1 and FPSO2.

\[ \Delta v_{arv}(t) = |v_{arv}(t) - v_{arv}(t - 1)| \] (14)

\[ v_{arv}(t) = \frac{1}{m \cdot D} \sum_{m=1}^{m} \sum_{d=1}^{D} v_{id} \]

Where m is the population size and D is the dimension of the problem.

We explain the proposed method in Fig. 4.1.

4.2. Combination of improved AFPSO and rough set theory for classification

Classification is an important issue in Artificial Intelligence. Proper features can effectively reduce computational cost and increase classification accuracy. For example if the number of features is n, then 2n different subsets with different length will be produced. Therefore considering all the features are not effective in classification. In other word, parts of the features are redundant and just increase the amount of computation or decrease the classification accuracy. Therefore we proposed a combination of improved AFPSO based on the Rough set theory called Rough Adaptive Fuzzy PSO (RAFPSO) for classification. The best position in the proposed algorithm is a subset with highest classification quality and minimum length. With more exploration and effective exploitation, we can have an optimized classification with the proposed approach. The proposed method illustrated in Fig. 4.2.

The merit of using rough theory in comparison to fuzzy theory is in having better feature reduction ability and doing the feature reduction with more accuracy. But fuzzy logic cannot eliminate redundant features and select the necessary feature [29].

Equal to N where N shows the total number of features. Similar to [18] the value 1 in each feature shows the relevant feature selection and the value 0 shows the not selected features.

4.2.1 Position

We consider each particle position a binary bit with length equal to N where N shows the total number of features. Similar to [18] the value 1 in each feature shows the relevant feature selection and the value 0 shows the not selected features.

4.2.2 Voracity

Particles voracity differs between 1 to vMax. Below we show how a particle’s trajectory toward the best particle over time. If a particle defined as P[i]=[0 0 1 1 0 1] and the best particle is defined as P[gbest]=[1 1 0 0 0 1], then he difference between each particle with the best particle is defined as P[i]-P[gbest]=[1 -1 1 0 0 0]. The value 1 in the chromosome string shows that the relevant feature is very important and thus if that feature is not selected, the classification accuracy will be decreased. In the same way, the value-1 shows that the feature is redundant and it is not proper to select that feature. If that feature is selected it only increase the computational cost and decreases the classification accuracy. The distance between two positions is gain from the differences of total number of the values 1 and total number of the values -1.
Figure 4.1 The flow chart of the proposed AFPSO method.

Figure 4.2 The flow chart of the proposed RAFPSO method.
4.2.3 Objective Function
We use the rough theory presented in [18].

\[ \text{Fitness} = \alpha \cdot \gamma_R(D) + \beta \cdot \left(1 - \frac{|B|}{|C|}\right) \quad (15) \]

Where the value of \( \alpha \) and \( \beta \) are the same as [18], \( \alpha \) is the weight which is used for rough theory and it is equal to 0.9. \( \gamma_R(D) \) is relevant to rough theory, \( C \) is the total features and \( B \) is the selected features.

5. SIMULATION
We implement the proposed approaches in Mat lab 2009. In the first approach, the design of the fuzzy inference system \( W \) for the AFPSO algorithm is the same as [30]. For each input and output of the fuzzy system, three membership function \( (S, M, L) \) where \( S \) is small, \( M \) is medium and \( L \) is large are defined. The 9 rules for \( W \) is presented in Fig. 6.1 evaluated using 4 UCI data sets, where the details are presented in table 3. We evaluate the proposed AFPSO method with the 4 mathematical benchmark functions presented in table 1. It is good to know that we use only the 2 dimension for the functions. We compared the result with the results presented in [27]. Then, RAFFPSO is evaluated using 4 UCI data sets, where the details are presented in table 3.

6. THE RESULT
6.1. Improved AFPSO
We evaluated the first proposed method with 4 mathematical benchmark functions. The results presented in table 2. The percentage of the error shows the high efficiency, and more precisions of the proposed method comparing the methods presented in [27]. However, due to this fact that the functions are not having the same global minimum, we normalized the result like in [27].

![Figure 6.1 Fuzzy Rules for W Production [30].](image)

### Table 1

<table>
<thead>
<tr>
<th>Number Function</th>
<th>Name Of Function</th>
<th>Mathematical Benchmark Functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( f(x, y) = x \sin(4x) + 10y \sin(2y) )</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>( f(x) = -a \cdot e^{-b \sum_{i=1}^{n} x_i^2} + a + e^{c \sum_{i=1}^{n} \cos(2x_i)} )</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>( f(x) = \sum_{i=1}^{n} \sin(x_i) \cdot \sin \left(\frac{x_i^2}{n}\right) )</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>( f(x) = \sum_{i=1}^{n} x_i^2 )</td>
<td></td>
</tr>
</tbody>
</table>

### Table 2

The results of mathematical Benchmark Functions

<table>
<thead>
<tr>
<th>AFPSO2</th>
<th>FPSO2[27]</th>
<th>AFPSO1</th>
<th>FPSO1 [27]</th>
<th>Pop Size</th>
<th>Num Of Iteration</th>
<th>Minimum</th>
<th>Range</th>
<th>function</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1448</td>
<td>0.2645</td>
<td>0.1732</td>
<td>0.2484</td>
<td>10</td>
<td>30</td>
<td>-18.5547</td>
<td>(0,10)</td>
<td>1</td>
</tr>
<tr>
<td>0.019</td>
<td>0.0157</td>
<td>0.0027</td>
<td>0.0039</td>
<td>10</td>
<td>30</td>
<td>0</td>
<td>(-2,2)</td>
<td>2</td>
</tr>
<tr>
<td>0.4487</td>
<td></td>
<td>0.6378</td>
<td></td>
<td>20</td>
<td>1000</td>
<td>-9.6601517</td>
<td>(0,( n ))^n</td>
<td>3</td>
</tr>
<tr>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>10</td>
<td>30</td>
<td>0</td>
<td>(-5.12,5.12)</td>
<td>4</td>
</tr>
</tbody>
</table>

### Table 3

Used Parameters.

<table>
<thead>
<tr>
<th>Data</th>
<th>Features</th>
<th>Instances</th>
<th>Pop Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breast issue</td>
<td>9</td>
<td>106</td>
<td>30</td>
</tr>
<tr>
<td>Breast Cancer</td>
<td>9</td>
<td>699</td>
<td>30</td>
</tr>
<tr>
<td>Hepatitis</td>
<td>19</td>
<td>155</td>
<td>30</td>
</tr>
<tr>
<td>Lung Cancer</td>
<td>56</td>
<td>32</td>
<td>30</td>
</tr>
</tbody>
</table>
Comparing zzy system forational hms in classification as a classifier in order to over
come X. Wang, et al., “Finding Minimal Rough Set Reducts

For the future works we aim
used data sets.

results in table 2 show the merits of the proposed method
set theory for finding subsets of optimized features. Present
method using an improved first proposed method and rough
accuracy on the several UCI data set, we proposed a hybrid
weight inertia [3] is used. In order to increase classification
between exploitation ad explorations, a fu
improving the performance and considering a trade of
parameters based on fuzzy set presented in [27]. Also for
apply two proposed methods that have dynamic adaptive

efficiency the problems of

In this paper, we use PSO a

7.2. The Results of the Combination of the Proposed
Improved AFPSO with Rough Set Theory on UCI Data
Set
We evaluate the second proposed approach on the 4 UCI data
sets. The details presented in table 3. Then we evaluate
the result with 3 other classification methods including FSUM
[31], FSLA [32] and FSMI [31]. The results in the table 4 are
derived from the table 5 in the paper [31].
The first and the second proposed systems on the Breast issue
data set comparing to FPSO1 and FPSO2 presented in [27]
have more accuracy. For the Breast Cancer Wisconsin data
set the second data set is improved about 55.06% comparing
to FPSO2.
The proposed method show good improvements in
classification of Hepatitis data set. Finally, for the Lung
Cancer the first proposed methods comparing to FSUM give
better results. But comparing to the second proposed method,
the accuracy is decreased about 3.14%. In table 2 and 4 the “-
“ is mean that there is no mathematical function or related
data set in the paper.
We use LEM2 similar to [34] for rule production and use
10foldcrossvalidation for accuracy estimation. The proposed
approaches run for 30 times and evaluated separately.

7. CONCLUSIONANDFUTURE WORKS
In this paper, we use PSO as a classifier in order to overcome
the problems of the current classifiers. However the PSO
efficiency is strictly related to its parameters. Hence, we
apply two proposed methods that have dynamic adaptive
parameters based on fuzzy set presented in [27]. Also for
improving the performance and considering a trade of
between exploitation ad explorations, a fuzzy system for
weight inertia [3] is used. In order to increase classification
accuracy on the several UCI data set, we proposed a hybrid
method using an improved first proposed method and rough
set theory for finding subsets of optimized features. Presented
results in table 2 show the merits of the proposed method
comparing to the first method presented in [27] and methods
presented in [31-33] although we have the missing data in the
used data sets.
For the future works we aim at,

- RAFPSO optimizing using game theory.
- applying AFPSO on functions with more than two
dimensions in comparison with other types of evolutionary algorithm.

REFERENCES
1. Z. Pawlak, "Rough sets," International Journal of
Computer & Information Sciences, vol. 11, pp. 341-356,
1982/10/01 1982.
2. Z. Pawlak, "Imprecise Categories, Approximations and
Rough Sets," in Rough Sets. vol. 9, ed: Springer
3. Z. Pawlak, "Rough set approach to knowledge-based
decision support," European Journal of Operational
4. R. W. Swiniarski and A. Skowron, "Rough set methods in
5. S. Satapathy, et al., "Rough set and teaching learning based
optimization technique for optimal features selection,"
Central European Journal of Computer Science, vol. 3,
pp. 27-42, 2013/03/01 2013.
Improved Reduct Generation in Information Systems,"
Knowledge and Information Systems, vol. 2, pp. 131-
146, 2000/06/01 2000.
7. A. Skowron and C. Rauszer, "The Discernibility Matrices
and Functions in Information Systems," in Intelligent
Decision Support. vol. 11, R. Słowiński, Ed., ed: Springer
8. X. Hu and N. Cercone, "LEARNING IN RELATIONAL
DATABASES: A ROUGH SET APPROACH,"
9. R. Susmaga, "Reducts and Constructs in Attribute
10. R. Susmaga, "Tree-Like Parallelization of Reduce and
Construct Computation," in Rough Sets and Current
Eds., ed: Springer Berlin Heidelberg, 2004, pp. 455-
464.
11. J. G. Bazan, et al., "Rough set algorithms in classification
problem," in Rough set methods and applications, ed:
12. X. Wang, et al., "Finding Minimal Rough Set Reducts
with Particle Swarm Optimization," in Rough Sets,
Fuzzy Sets, Data Mining, and Granular Computing.
vol.

May-June
2070

ISSN 1013-5316; CODEN: SINTE 8


May-June