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

Hamid Tabatabaee[a,*] ,Parisa Molavi[b], Maryam Baratzadeh[b], NasibehRadyRaz[c]

[a,*]Young Researchers and Elite Club, Quchan Branch, Islamic Azad University, Quchan, Iran.
[b]Department of Computer engineering, Mashhad branch, Islamic Azad University, Mashhad, Iran.
[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

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

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 date set becoming 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 in to 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.

Fuzzy logic or three values logic based on the fuzzy logic theory is presented by Professor Lotfi Zadeh in 1965 [22]. Fuzzy logic theory can model linguistic knowledge by using if… then rules [23].

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 objects. A is a non-empty and finite set of features. $\forall a \in A$is a function that defined $f_a: U \to V_a$ in which V_a is a set of values of a. If $P \subseteq A$ then there is an equivalence relation as follows:

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

U part which is produced by IND (P) means U/P. If $(x, y) \in IND(P)$, 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 ($\underline{p}$x)p and the upper approximation is ($\overline{p}$x)p and they will be defined as follows:

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

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

If $\mathcal{Q} \subseteq A$ and P is the equivalence relation, then the positive, negative and boundary regions will be defined as follows:

$$POS_P(\mathcal{Q}) = \bigcup_{X \in U/\mathcal{Q}} \underline{P}X \quad (4)$$

$$NEG_P(\mathcal{Q}) = \bigcup_{X \in U/\mathcal{Q}} \overline{P}X \quad (5)$$

$$BND_P(\mathcal{Q}) = \bigcup_{X \in U/\mathcal{Q}} \overline{P}X - \bigcup_{X \in U/\mathcal{Q}} \underline{P}X \quad (6)$$

The positive region from the U/P section is in relation to set of p which is shown by $POS_P(\mathcal{Q})$ 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 nun-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 $IND(P) \subseteq IND(\mathcal{Q})$. 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 $P \Rightarrow \mathcal{Q}_k$and equal to:

$$k = \gamma_P(\mathcal{Q}) = \frac{|POS_P(\mathcal{Q})|}{|U|} \quad (7)$$

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]:

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

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

$$Re\,d_{min} = \{R \in Re\,d \mid \vee\ R' \in Re\,d\,, ||R| \leq |R'|\} \quad (9)$$

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

$$Core(C) = \bigcap Re\,d \quad (10)$$

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).

$$v_{id} = w * v_{id} + c_1 * \mathrm{Rand}() * (p_{id} - x_{id}) + c_2 * \mathrm{Rand}() * (p_{gd} - x_{id}) \quad (11)$$

$$x_{id} = x_{id} + v_{id} \quad (12)$$

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. THEPROPOSED 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 inertiaprepare 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.

Iteration:

$$\text{Iteration=current iteration} \quad (13)$$

The average weight ratio:

$$\begin{cases} \Delta v_{arv}(t) = |v_{arv}(t) - v_{arv}(t-1)| \\ v_{arv}(t) = \dfrac{1}{m.D}\sum_{m}\sum_{D} v_{id} \end{cases} \quad (14)$$

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 Pi=[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 Pi-P_{gbest}=[-1 -1 1 1 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 decrees 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.

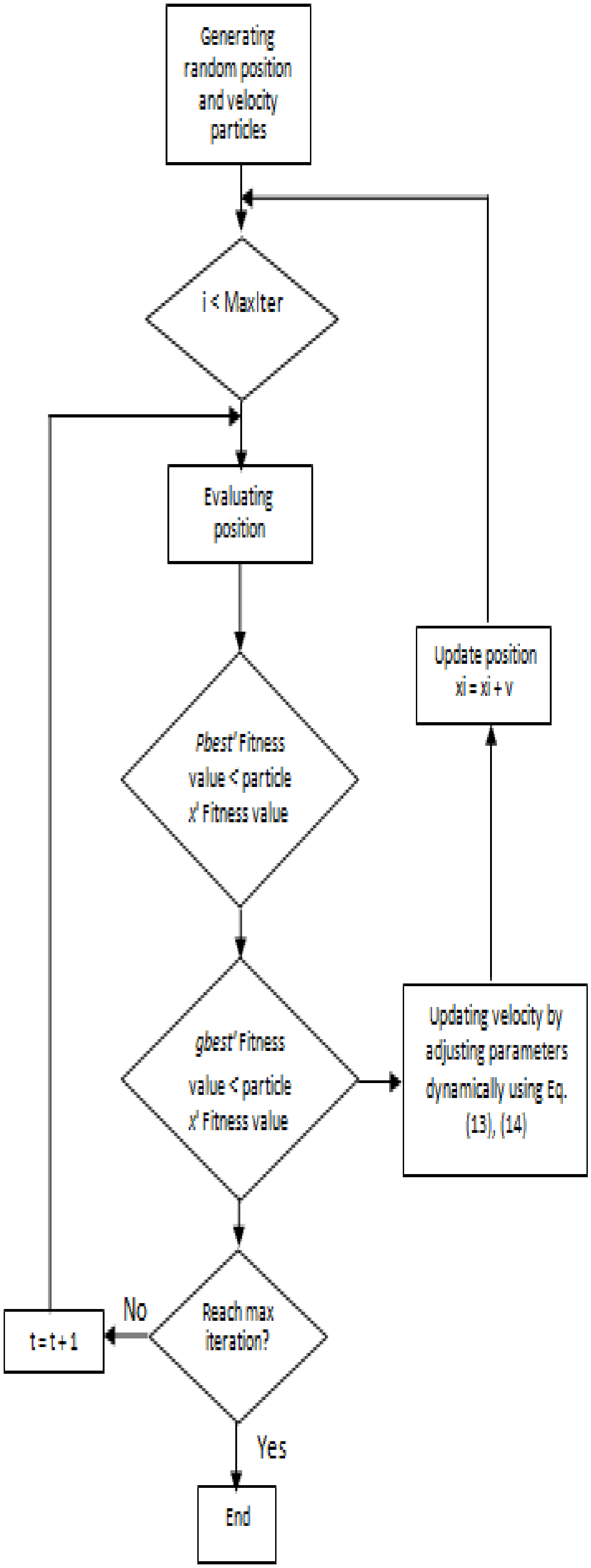

Figure 4.1 The flow chart of the proposed AFPSO method

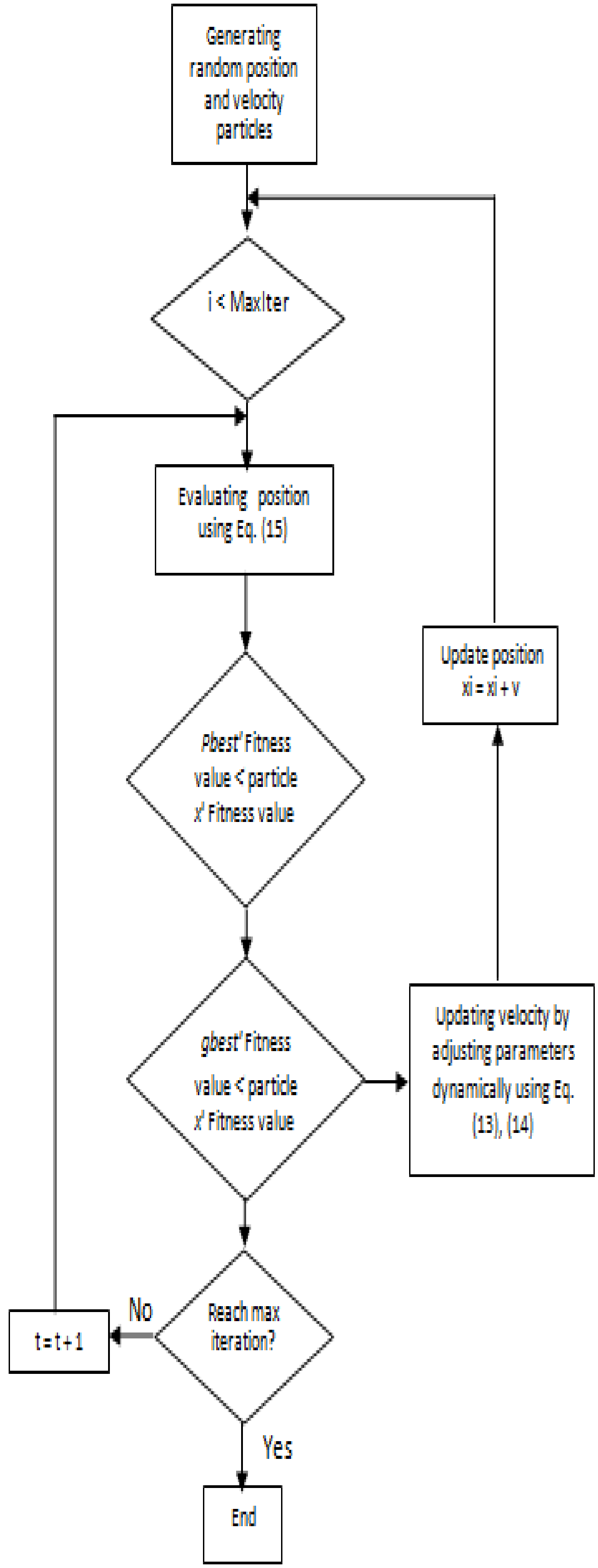

Figure 4.2 The flow chart of the proposed RAFPSO method

4.2.3 Objective Function

We use the rough theory presented in [18].

$$\mathrm{Fitness} = \alpha * \gamma_B(D) + \beta * \left(1 - \frac{|B|}{|C|}\right) \quad (15)$$

Where the value of α and β are the same as [18]. α is the weight which is used for rough theory and it is equal to 0.9. βWhich is used for feature length elimination is equal to 0.1. $\gamma_R(D)$ 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. 6.1evaluated 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, RAFPSO 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].

1. If (iter is S) and (V_{avr} is S) Then (w is L)
2. If (iter is M) and (V_{avr} is S) Then (w is L)
3. If (iter is L) and (V_{avr} is S) Then (w is M)
4. If (iter is S) and (V_{avr} is M) Then (w is L)
5. If (iter is M) and (V_{avr} is M) Then (w is M)
6. If (iter is L) and (V_{avr} is M) Then (w is S)
7. If (iter is S) and (V_{avr} is L) Then (w is M)
8. If (iter is M) and (V_{avr} is L) Then (w is S)
9. If (iter is L) and (V_{avr} is L) Then (w is S)

Figure 6.1 Fuzzy Rules for W Production [30].

Table 1

Mathematical Benchmark Functions

Number OF Function	Name OF Function
1	$f(x,y) = x\sin(4x) + 10y\sin(2y)$
2	$f(x) = -a.e^{-b\sqrt{\frac{\sum_{i=1}^{n} x^2}{n}}} - e^{\sum_{i=1}^{n}\cos(2x_i)} + a + e^1$
3	$f(x) = \sum_{t=1}^{n} \sin(x_1).\left\|\sin\left(\frac{x_t^2}{n}\right)\right\|^{20}$
4	$f(x) = \sum_{i=1}^{n} x_i^2$

Table 2

The results of mathematical Benchmark Functions

AFPSO2	FPSO2[27]	AFPSO1	FPSO1 [27]	Pop Size	Num OF Iteration	Minimum	Range	function
0.1448	0.2645	0.1732	0.2484	10	30	−18.5547	(0,10)	**1**
0.019	0.0157	0.0027	0.0039	10	30	0	(−2,2)	**2**
0.4487	–	0.6378	–	20	1000	−9.6601517	$(0,\pi)^n$	**3**
0.0000	0.0000	0.0000	0.0000	10	30	0	(−5.12,5.12)	**4**

Table 3

Used Parameters.

Data	Features	Instances	Pop Size
Breast issue	9	106	**30**
Breast Cancer Wisconsin	9	699	**30**
Hepatitis	19	155	**30**
Lung Cancer	56	32	**30**

Table 4

Results on the UCI Data set.

Data	RAFPSO1	RAFPSO 2	FPSO1[27]	FPSO2[27]	FSUM[31]	FSLA[32]	FSMI[33]
Breast tissue	**94.73**	**92.54**	92.86	87.62	-	-	-
Breast Cancer Wisconsin	**96.32**	**90.43**	94.07	34.37	85.15 ± 1.64	82.66 ± 2.01	$\mathbf{77.90 \pm 1.83}$
Hepatitis	**95.92**	**96.74**	-	-	82.99 ± 1.09	81.02 ± 1.26	$\mathbf{80.65 \pm 1.41}$
Lung Cancer	**93.03**	**89.26**	-	-	89.13 ± 0.76	85.25 ± 0.30	$\mathbf{82.71 \pm 0.55}$

6.2. 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,

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

REFRENCES

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 Netherlands, 1991, pp. 9-32.
3. Z. Pawlak, "Rough set approach to knowledge-based decision support," European Journal of Operational Research, vol. 99, pp. 48-57, 1997.
4. R. W. Swiniarski and A. Skowron, "Rough set methods in feature selection and recognition," Pattern Recogn.Lett., vol. 24, pp. 833-849, 2003.
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.
6. J. A. Starzyk, et al., "A Mathematical Foundation for 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 Netherlands, 1992, pp. 331-362.
8. X. Hu and N. Cercone, "LEARNING IN RELATIONAL DATABASES: A ROUGH SET APPROACH," Computational Intelligence, vol. 11, pp. 323-338, 1995.
9. R. Susmaga, "Reducts and Constructs in Attribute Reduction," Fundam. Inf., vol. 61, pp. 159-181, 2004.
10 .R. Susmaga, "Tree-Like Parallelization of Reduct and Construct Computation," in Rough Sets and Current Trends in Computing. vol. 3066, S. Tsumoto, et al., 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: Physica-Verlag GmbH, 2000, pp. 49-88.
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.

3641, D. Ślęzak, et al., Eds., ed: Springer Berlin Heidelberg, 2005, pp. 451-460.
13. L.-Y. Zhai, et al., "Feature extraction using rough set theory and genetic algorithms—an application for the simplification of product quality evaluation," Computers & Industrial Engineering, vol. 43, pp. 661-676, 2002.
14. A. Skowron, et al., "RSES 2.2 user's guide," Institute of Mathematics, Warsaw University, Warsaw, Poland, 2005.
15. A. Skowron, et al., "A Hierarchical Approach to Multimodal Classification," in Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing. vol. 3642, D. Ślęzak, et al., Eds., ed: Springer Berlin Heidelberg, 2005, pp. 119-127.
16. R. Jensen and Q. Shen, "Finding rough set reducts with ant colony optimization," in Proceedings of the 2003 UK workshop on computational intelligence, 2003.
17. S. Xia, et al., "A Web Document Classification Method Based on Shuffled Frog Leaping Algorithm," in Genetic and Evolutionary Computing, 2008. WGEC '08. Second International Conference on, 2008, pp. 205-208.
18. S.-Y. Jing, "A hybrid genetic algorithm for feature subset selection in rough set theory," Soft Computing, vol. 18, pp. 1373-1382, 2014/07/01 2014.
19. K. Neshatian and M. Zhang, "Pareto front feature selection: using genetic programming to explore feature space," presented at the Proceedings of the 11th Annual conference on Genetic and evolutionary computation, Montreal, Québec, Canada, 2009.
20. R. Jensen, "Performing Feature Selection with ACO," in Swarm Intelligence in Data Mining. vol. 34, A. Abraham, et al., Eds., ed: Springer Berlin Heidelberg, 2006, pp. 45-73.
21. A. Unler and A. Murat, "A discrete particle swarm optimization method for feature selection in binary classification problems," European Journal of Operational Research, vol. 206, pp. 528-539, 2010.
22. L. A. Zadeh, "Fuzzy sets," Information and control, vol. 8, pp. 338-353, 1965.
23. C. Li and T. Wu, "Adaptive fuzzy approach to function approximation with PSO and RLSE," Expert Systems with Applications, vol. 38, pp. 13266-13273, 2011.
24. J. Kennedy and W. M. Spears, "Matching algorithms to problems: an experimental test of the particle swarm and some genetic algorithms on the multimodal problem generator," in Proceedings of the IEEE international conference on evolutionary computation, 1998, pp. 78-83.
25. R. C. Eberhart and J. Kennedy, "A new optimizer using particle swarm theory," in Proceedings of the sixth international and human science, 1995, pp. 39-43.
26. S. Intelligence, "Particle swarm optimization," 2007.
27. P. Melin, et al., "Optimal design of fuzzy classification systems using PSO with dynamic parameter adaptation through fuzzy logic," Expert Systems with Applications, vol. 40, pp. 3196-3206, 2013.
28. E. roghanian, et al, "using Rough sets Theory to reduce data bases redundant data," 1st National Conference on Applications of Intelligent Systems(Soft Computing) in Science and Technology, Islamic Azad University, Quchan Branch, 4-5, March, 2013, (in Persian).
29. R. Kokabi, et al, "Optimization of Black Box 0-1 problems with hybrid Scatter search and Rough sets," in 9th international industrial Engineering Conference 2013, IIEC09, (in Persian).
30. P. Yadmellat, et al., "A new fuzzy inertia weight particle swarm optimization," in Computational Intelligence and Natural Computing, 2009. CINC'09. International Conference on, 2009, pp. 507-510.
31. W. Qian, et al., "Feature Selection using Compact Discernibility Matrix-based Approach in Dynamic Incomplete Decision System," Journal of Information Science and Engineering, 2014.
32. X. Jin, et al., "Machine Learning Techniques and Chi-Square Feature Selection for Cancer Classification Using SAGE Gene Expression Profiles," in Data Mining for Biomedical Applications. vol. 3916, J. Li, et al., Eds., ed: Springer Berlin Heidelberg, 2006, pp. 106-115.
33. D. Huang and T. W. S. Chow, "Effective feature selection scheme using mutual information," Neurocomputing, vol. 63, pp. 325-343, 2005.
34. J. Stefanowski, "On rough set based approaches to induction of decision rules," Rough sets in knowledge discovery, vol. 1, pp. 500-529, 1998.