Method of Generating Contexts Based on Self-adaptive Differential Particle Swarm Using Local Topology for Multimodal Optimization in the Case of Multigranulation

Dianne Arias[1], Yaima Filiberto[1], Rafael Bello[2]

[1] University of Camaguey, Department of Computer Science, Cuba
dianne.arias@reduc.edu.cu, yaima.filiberto@reduc.edu.cu
[2] University of Las Villas, Department of Computer Science, Cuba
rbellop@uclv.edu.cu

Abstract. Several extensions of the rough set theory have been proposed in terms of various requirements an example of this is multi-granulation, where several separability relationships are used to obtain different granulations of the universe. In this approach we begin from the existence of different contexts or subsets of features to characterize the objects of the universes. Multigranulation is very desirable in many real applications and has been used to develop various learning techniques. It is usually part of the existence of these contexts. In this paper a method for the generation of contexts is proposed from the construction of relations of similarity with the use of multimodal PSO, which is used in order to obtain multiple maximums global (gbest) from which the set of contexts is created and the number of characteristics per context. The proposed method has been tested in improving the efficiency of the k-nn method, using different data sets.

Keywords: multigranulation, similarity quality measure, multimodal PSO.

1 Introduction

Rough set theory, originated by Pawlak [1–3], has become a well-established mechanism for uncertainty management in a wide variety of applications related to artificial intelligence [4–7]. One of the strengths of rough set theory is that all its parameters are obtained from the given data. This can be seen in the following paragraph from [8]: "The numerical value of imprecision is not pre-assumed, as it is in probability theory or fuzzy sets but is calculated on the basis of approximations which are the fundamental concepts used to express imprecision of knowledge". In other words, instead of using, the rough set data analysis (RSDA) utilizes solely the granularity structure of the given data, expressed as classes of suitable equivalence relations.

In the past 10 years, several extensions of the rough set model have been proposed in terms of various requirements, such as the variable precision rough

set (VPRS) model [9], the rough set model based on tolerance relation [10], the Bayesian rough set model [11], the fuzzy rough set model and the rough fuzzy set model [12].

In many circumstances, we often need to describe concurrently a target concept through multi binary relations according to a user's requirements or targets of problem solving, for that another extension of the RST is to use more than one separability relationship to perform the granulation of the universe, which is known as multigranulation [13, 14]. In this case, from the set of predictive features A, two or more subsets $A_1, ..., A_k$, A_i $Am \subseteq A$, are formed of features that allow defining the separability relation. These subsets of features are called contexts [15]. Based on this multigranulation approach, different techniques for the discovery of knowledge have been formulated.

2   Multigranulation in the Rough Set Theory

Qian *et al.* [13] proposed multigranulation rough set (MGRS) in complete information system to more widely apply rough set theory in practical applications, in which lower/upper approximations are approximated by granular structures induced by multi binary relations. The multigranulation rough set is different from Pawlak's rough set model because the latter is constructed on the basis of a family of indiscernibility relations instead of single indiscernibility relation. In optimistic multigranulation rough set approach, the word 'optimistic' is used to express the idea that in multi independent granular structures, we need only at least one granular structure to satisfy with the inclusion condition between equivalence class and the approximated target. The upper approximation of optimistic multigranulation rough set is defined by the complement of the lower approximation [37].

From the point of view of the applications of the RST, the multigranulation in the RST is very desirable in many real applications, such as analysis of data from multiple sources, discovery of knowledge to from data with large dimensions and distributive information systems. Since Qian in 2006 proposed multigranulation in the RST, the theoretical framework has been widely enriched, and many extensions of these models have been proposed and studied [35, 36]. In the mutigranulation rough set theory, each of various binary relation determines a corresponding information granulation, which largely impacts the commonality between each of the granulations and the fusion among all granulations.

In their papers, Qian *et al.* said that the MGRS are useful in the following cases:

1. We cannot perform the intersection operations between their quotient sets and the target concept cannot be approximated by using $U/(P \bigcup Q)$ which is called a single granulation in those papers.

2. In the process of some decision making, the decision or the view of each of decision makers may be independent for the same project (or a sample, object and element) in the universe. In this situation, the intersection operations between any two quotient sets will be redundant for decision making.

3. Extract decision rules from distributive information systems and groups of intelligent agents through using rough setapproaches

Since then,many researchers have extended the classical MGRS by using various generalized binary relations.

3 Self-Adaptive Differential Particle Swarm Using a Local Topology for Multimodal Optimization

Particle Swarm Optimization is an effective and robust non-direct global-search method for solving challenging continuous optimization problems. The PSO meta-heuristic involves a set of particles known as swarm which explore the search space trying to locate promising regions [32]. Therefore, particles are interpreted as solutions for the optimization problem and they are represented as points in n-dimensional search space. In the case of standard PSO, each particle X_i moves through the space using its own velocity V_i, a local memory of the best position it has obtained P_i and knowledge of the best solution G found in its neighborhood. Equations 1 and 2 show how to update the particles position based on the mentioned components:

$$\vartheta_i(t + 1) = \alpha * \vartheta_i(t) + U(0, \varphi1)(pBest(t) - x_i)$$
$$+ U(0, \varphi2(gBest(t) - x_i), \tag{1}$$

$$x_i(t + 1) = x_i(t)\vartheta_i(t + 1). \tag{2}$$

In last decades Evolutionary and Swarm Intelligence algorithms have become an important improvement for both discrete and real-parameter optimization. Without niching [34] strategies they converge to a single optimum, even in multimodal search spaces where numerous global or local solutions exist. However, most real-life problems are characterized by multimodal functions. In literature several niching approaches have been proposed for computing multiple optima simultaneously, though most of them require some user-specified parameters that should be estimated in advanced (i.e. additional knowledge about problem domain is required) [34].

Then, multimodal optimization methods try to discover and maintain multiple subpopulations in a single run, where each niche corresponds to a specific peak of the fitness landscape (ideally one species per optimum). They have been developed to reduce the undesirable effects of genetic drift. In few words, niching strategies should be able to preserve the diversity in the artificial population, allowing individuals parallel convergence toward different solutions. As well, niching methods are useful to avoid stagnation or premature convergence states in global optimization problems where many sub-optimal solutions exist; offering an escaping alternative from local optima [34].

When multimodal problems are solved, the main advantage of the lbest model appears to lie in its slower convergence rate relative to the gbest model, allowing concurrently discovering several optima. Ironically, it is the slightly interaction among particles that is most responsible of the poor performance of the PSO

Dianne Arias, Yaima Filiberto, Rafael Bello

based algorithms using a Ring Topology. To improve the search capability of such models a novel Differential Operator is introduced. This operator is straightfor- wardly inspired on the well-known differential strategy DE/current-to-rand/1 without crossover [33]. Therefore, as first step, we design a mutation operator as illustrate following equation 3:

$$\widetilde{x}_i(t+1) = pBest_i(t) + F * (pBest_{r1}(t) - pBest_{r2}(t)), \tag{3}$$

where $pBest_i(t)$ denotes the personal best position of current individual, $pBest_{r1}(t)$ and $pBest_{r2}(t)$ are the global best record achieved by two randomly selected swarm particles.

Next, a selection operation takes place, where $\widetilde{x}_i$ is accepted as current particle position if it improves the search procedures, respect to the solution generated by the PSO rules; otherwise the mutant is rejected (See in equation 4) [34]:

$$x_i^{t+1} = \begin{cases} x_i^{t+1}, \ sif \left(\widetilde{x}_i^{\,t+1}\right) \leq f\left(P_i^{t+1}\right), \\ x_i^{t+1}, \ sif \left(\widetilde{x}_i^{\,t+1}\right) > f\left(P_i^{t+1}\right). \end{cases} \tag{4}$$

Following a similar reasoning of the conventional clearing, it's used a novel diversity procedure: Heuristic Clearing. It is able to preserve the swarm diversity in lbest PSO algorithms using a Ring Topology based topology, and it does not need to be specified any niche parameter. To do that, this operator only takes into account optimal particles (See equation 5) [34]:

$$\mid f\left(P_i\right) - F^* \mid < \varepsilon. \tag{5}$$

In the following section we propose a method to generate contexts using multigranulation based on the rough set theory and multimodal PSO.

4 Method of Generating Contexts Based on Self-adaptive Differential Particle Swarm Using Local Topology for Multimodal Optimization in the Case of Multigranulation

Be a decision system $SD = A \bigcup d$ where the domain of the characteristic in $A \bigcup d$ may be discrete or continuous values, from which calculate the features weights using the PSOMulti+RST+MG method, which is a modification of PSO + RST [17]. In this case PSO Multimodal is used in order to obtain multiple maximums global (gbest) from which the set of contexts is created and the number of characteristics per context, then weights are ordered by contexts and those with a weight greater than the mean value are selected of the weight for that context. Finally the same contexts are remove. The algorithm is described below.

Algorithm 1 Pseudocode for PSOMulti+RST+MG algorithm.

---

    1. Calculate the weights (w) of features using PSOMulti+RST method.
    2. Generate set of contexts using GBest(Φn)
        $C = \Phi n$
    3. For each context C_i:
        Order W_i
        Select $W_{ij} \in W_{imax} \iff \{W_{ij} > mean(W_i)\}$
    4. Select de different context
        $\forall C_i, C_j | C_i, C_j \exists C \wedge C_i \neq C_j$

Algorithm PSOMULTI+RST+MG

Step 1: Initialize a population of particles with random positions and velocities in a D-dimensional space.

Step 2: For each particle, evaluate the quality measure of similarity using expression 6, in D Variables.

$$\max \rightarrow \left\{ \frac{\sum\limits_{\forall x \epsilon U} \varphi(x)}{|U|} \right\}. \tag{6}$$

    Step 3: Compare the quality measure of the current similarity of each particle with the quality measurement of the similarity of your previous best position $pBest$. If the current value is better than that of $pBest$, then assign to $pBest$ the current value, and $pBest = x_i$, that is, the current location becomes the best one so far.

    Step 4: Identify the particle in the neighborhood with the highest value for the quality of similarity measure and assign its index to the variable $gBest$ and assign the best value of the quality measure of similarity to m.

    Step 5: Adjust the speed and position of the particle according to equations 1, 2 and 3 (for each dimension).

    Step 6: Verify if the stop criterion is met (maximum number of iterations or if it takes five iterations without improving the quality measure of the global similarity (m)), if not, go to Step2.

5 Experimental Results

For this study we used data sets from the UCI repository [38] (*iris*, *schizo*, *soybeansmall*, *hepatitis*, *dermatology*, *sonar*, *lungcancer*, *biomed*, *glass*, *analcaa−bankruptcy*, *segment*). It is used to calculate the weights for KNN [23] with $k = 1$ the proposed method. The training and test sets were obtained, taking 75 percent of the cases for the first and 25 percent for the second, in a totally randomized manner. Following this principle of random selection the process was repeated ten times and ten training sets and ten test sets were obtained for each

data set, in order to apply cross validation [24] for a better validation of the results.

The parameters used in the experimentation, for the method $PSOMulti + RST + MG$ were: $TB = 40$, $NI = 100$, $ce1 = ce2 = 2$ and the values of $e1$ and $e2$ for the function of similarity between attributes and for the function of similarity for the decision attribute were between 0.65-0.83 and 1.0, F=0.1. The stop condition is: when 100 iterations are reached or when in five iterations the fitness value does not improve (measure quality similarity quality).

It is used as a KNN classifier with $K = 1$ to make a comparison of the results obtained after the creation of the proposed method's contexts $PSOMULTI + RST+MG$ with algorithms $AdaBoostM1[30]$, $RandomSubSpace$ [31] and $Bagging$ [29], implemented in the WEKA tool and using the KNN as a classifier, in all cases.

For the statistical analysis of the results, the hypothesis testing techniques were used [25]. For multiple comparisons, the Friedman and Iman-Davenport tests [26] are used to detect statistically significant differences between a groups of results. The Holm test [27] is also used in order to find significantly higher algorithms.

These tests are suggested in the studies presented in [24], which states that the use of these tests is highly recommended for the validation of results in the field of automated learning. In the statistical processing of all the experimental results, the KEEL was used [28].

Table 1 shows the description of the data sets used in the experimentation, as well as the contexts obtained by the proposed method (column 4) and the number of average features for each context (column 5).

Table 2 shows the results of evaluating the contexts with the KNN method for $K = 1$ (Knn algorithm $PSOMULTI + RST + MG$), as well as the results of the $AdaBoostM1$, $RandomSubSpace$ and $Bagging$ algorithms, as you can observe the proposed method obtains better results than the rest.

Table 1. Datasets.

Datasets	Instances	Feature	Contexts	Features average X Contexts
iris	150	4	5	2.4
schizo	104	14	14	11.36
soybean-small	47	35	29	33.35
hepatitis	155	19	29	15.55
Dermatology	358	34	14	32.43
sonar	208	60	34	30.77
lung-cancer	32	56	10	44.6
biomed	194	8	14	6.43
glass	214	9	14	7.36
analcaa-bankruptcy	50	6	10	4.5
segment	2310	19	13	16.62

Table 2. Experimental results for KNN with $K=1$.

Dataset	AdaBoostM1	RandomSubSpace	Bagging	PSOMulti+RST+MG
iris	94.67⊖	91.33⊖	94.67⊖	95.8⊕
schizo	59.45⊖	60.45⊖	58.64⊖	62.8⊕
soybean-small	100⊙	97.5⊖	100⊙	100⊙
hepatitis	80.71⊖	82.63⊖	80.75⊖	82.8⊕
Dermatology	92.99⊖	95.25⊖	92.71⊖	96.1⊕
sonar	86.45⊖	87.55⊖	84.14⊖	89.5⊕
lung-cancer	65⊖	68.33⊖	68.33⊖	81.6⊕
biomed	89.71⊖	90.76⊖	89.71⊖	97.8⊕
glass	72.06⊖	77.98⊖	71.58⊖	78⊕
analcaa-bankruptcy	84⊖	86⊖	88⊙	88⊙
segment	97.23⊖	97.06⊖	97.01⊖	97.4⊕

Thus, ⊕ indicates that the accuracy is significantly better when $PSOMULTI+RST+MG$ method is used, ⊖ signifies that the accuracy is significantly worse and ⊙ signifies that there is no significant differences.

The Holm test was applied, with respect to the general accuracy of the KNN, and it is corroborated that the results are significantly higher when the contexts obtained by the $PSOMULTI+RST+MG$ method are used. Tables 3 and 4 show the results of the statistical tests related to this result.

P-values obtained in by applying post hoc methods over the results of Friedman procedure. Average ranks obtained by each method in the Friedman test.

Table 3. Average Rankings of the algorithms (Friedman).

Algorithm	Ranking
AdaBoostM1	3.0909
RandomSubSpace	2.5909
Bagging	3.1818
PSOMulti+RST+MG	1.1364

Iman and Davenport statistic (distributed according to F-distribution with 3 and 30 degrees of freedom): 11.568627.
P-value computed by Iman and Daveport Test: 0.000033099305.

Table 4. Post Hoc comparison Table for $\alpha=0.05$ (FRIEDMAN).

i	Algorithm	$z = (R_0 - R_i)/SE$	p	Holm
3	Bagging	3.715753	0.000203	0.016667
2	AdaBoostM1	3.550608	0.000384	0.025
1	RandomSubSpace	2.642313	0.008234	0.05

6 Conclusion

In this paper a new method of generating contexts based on similarity relationships for multigranulation using Self-adaptive Differential Particle Swarm using Local Topology for Multimodal Optimization is proposed. The main contribution is the construction of similarity relations based on the quality of similarity measure of Rough Sets Theory as a function of membership to build contexts for multigranulation. This measure calculates the degree of similarity in a decision system in which the feature may have discrete or continuous values. The contexts obtained were evaluated in international databases with the k-NN. The results achieved were significantly superior to the compared methods, which shows the effectiveness of the proposed method.

References

1. Pawlak, Z.: Rough Sets. international journal of Computer and Information Sciences 11, pp. 341–356 (1982)
2. Komorowski, J., Pawlak, Z.: Rough Sets: A tutorial. Rough Fuzzy Hybridization: A new trend in decision-making. Springer, pp. 3–98 (1999)
3. Greco, S.: Rough sets theory for multicriteria decision analysis. European Journal of Operational Research 129, pp. 1–47, (2001)
4. Duntsch, I., Gediga, G.: Uncertainty measures of rough set prediction. Artificial Intelligence 106, pp. 109–137 (1998)
5. Jensen, R., Shen, Q.: Fuzzy-rough sets assisted attribute selection. IEEE Transactions on Fuzzy Systems 15(1), pp. 73–89 (2007)
6. Jeon, G., Kim, D., Jeong, J.: Rough sets attributes reduction based expert system in interlaced video sequences. IEEE Transactions on Consumer Electronics 52(4), pp. 1348–1355 (2006)
7. Liang, J.Y., Dang, C.Y., Chin, K.S., Yam Richard, C.M.: A new method for measuring uncertainty and fuzziness in rough set theory. International Journal of General Systems 31(4), pp. 331–342 (2002)
8. Pawlak, Z.: Rough Sets: Theoretical Aspects of Reasoning about Data. System Theory, Knowledge Engineering and Problem Solving, vol. 9, Kluwer, Dordrecht (1991)
9. Ziarko, W.: Variable precision rough sets model. Journal of Computer System Science 46(1), pp. 39–59 (1993)
10. Skowron, A., Stepaniuk, J.: Tolerance approximation spaces. Fundamenta Informaticae 27 (2-3), pp. 245–253 (1996)
11. Slezak, D., Ziarko, W.: The investigation of the Bayesian rough set model. International Journal of Approximate Reasoning 40, pp. 81–91 (2005)
12. Dubois, D., Prade, H.: Rough fuzzy sets and fuzzy rough sets. International Journal of General Systems 17, pp. 191–209 (1990)
13. Qian, Y.H., Liang, J.Y., Yao, Y.Y., Dang, C.Y.: MGRS: a multi-granulation rough set. Information Sciences 180, pp. 949–970 (2010)
14. Lin, G., Liang, J., Qian, Y.: Multigranulation rough sets: From partition to covering. Information Sciences 241, pp. 101–118 (2013)
15. Intan, R., Mukaidono, M.: Multi-rough Sets Based on Multi-contexts of Attributes. Lectures Notes on Artificial Intelligence 2639, pp. 279–282 (2003)

16. Filiberto, Y., Bello, R., Caballero, Y., Larrua, R.: A method to build similarity relations into extended Rough Set Theory. In: 10th International Conference on Intelligent Systems and Applications ISDA 2010. Cairo, Egypt. IEEE Catalog Number CFP 10384 CDR (2010)
17. Filiberto, Y., Bello, R., Caballero, Y., Larrua, R.: Using PSO and RST to predict the resistant capacity of connections in composite structures. Studies in Computational Intelligence. Springer Berlin / Heidel-berg. Volume 284/2010 (2010)
18. Qian, Y., Liang, J.: Combination Entropy and Combination Granulation in Incomplete Information System (2006)
19. Qian, Y., Liang, J., Dang, C.: Knowledge structure, knowledge granulation and knowledge distance in a knowledge base. International Journal of Approximate Reasoning (2009)
20. Epitropakis, M.G., Plagianakos, V.P., Vrahatis, M.N.: Finding multiple global optima exploiting differential evolutions niching capability. In: IEEE Symposium on Differential Evolution, pp. 1–8 (2011)
21. Bird, S. and Li, X.: Adaptively choosing niching parameters in a PSO. In: Proc. of the 8th Annual Conference on Genetic and Evolutionary Computation GECCO 06, pp. 3–10 (2006)
22. Liu, Y., Ling, X., Shi, Z., Mingwei, L.V., Fang, J., Zhang, L.: A survey on Particle Swarm Optimization algorithms for multimodal function optimization. Journal of Software 6(12), pp. 2449–2455 (2011)
23. Zaldivar, J.M.: Estudio e incorporación de nuevas funcionalidades al k-NN Workshop v1.0. Computación. Camagüey, Universidad de Camagüey (2008)
24. Demsar, J.: Statistical comparisons of classifiers over multiple data sets. Journal of Machine Learning Research (7), pp. 1–30 (2006)
25. Sheskin, D.: Handbook of parametric and nonparametric statistical procedures. Chapman and Hall/CRC (2007)
26. Iman, R., Davenport, J.: Approximations of the critical region of the friedman statistic. Communications in Statistics, Part A Theory Methods 9, pp. 571–595 (1980)
27. Holm, S.: A simple sequentially rejective multiple test procedure. Journal of Statistics 6, pp. 65–70 (1979)
28. Alcalá, J., Fernández, A.: KEEL Data-Mining Software Tool: Data Set Repository, Integration of Algorithms and Experimental Analysis Framework. Journal of Multiple-Valued Logic and Soft Computing (2010)
29. Breiman, L.: Bagging predictors. Machine learning 24(2), pp. 123–140 (1996)
30. Freund, Y., Schapire, R.E.: Experiments with a new boosting algorithm, 148–156. View Article PubMed/NCBI Google Scholar (1996)
31. Kam, T.H.: The random subspace method for constructing decision forests. IEEE transactions on pattern analysis and machine intelligence 20(8), pp. 832–844 (1998)
32. Kennedy, J., Eberhart, R.: Particle Swarm Optimization. In: Proc. of the IEEE Int. Conference on Neural Networks, pp. 1942–1948 (1995)
33. Storn, R., Price, K.: Differential Evolution - A simple and efficient heuristic for global optimization over continuous spaces. J. of Global Optimization, pp. 341–359 (1997)
34. Napoles, G., Grau, I., Bello, R., Falcon, R., Abraham, A.: Self-adaptive differential particle swarm using a ring topology for multimodal optimization. In: Intelligent Systems Design and Applications (ISDA), 2013 13th International Conference on, pp. 35–40 (2013)

35. Qian, Y.H., Dang, C.Y., Liang, J.Y.: MGRS in incomplete information systems. In: Proceedings of 2007 IEEE Conference on Granular Computing, pp. 163–168 (2007)
36. Qian, Y.H., Liang, J.Y., Dang, C.Y.: Incomplete multigranulation rough set. IEEE Transactions on Systems, Man and Cybernetics, Part A 20, pp. 420–431 (2010)
37. Qian, Y., Zhang, H., Sang, Y., Liang, J.: Multigranulation decision-theoretic rough sets. International Journal of Approximate Reasoning 55(1), pp. 225–237 (2014)
38. Asuncion, A., Newman, D.: UCI machine learning repository. A study of the behaviour of several methods for balancing machine learning training data. SIGKDD Explorations 6(1), pp. 20–29 (2007)