

Hybrid Rough Sets and Particle Swarm Optimization Application in Data Mining

D. S. Morshedy
Department of Mathematics
and
Computer science,
Faculty of Science, port said
University, Egypt

W. A. Awad
Department of Mathematics
and
Computer science,
Faculty of Science, port said
University, Egypt

M. M. Elgenidy
Department of Mathematics
and
Computer science,
Faculty of Science, port said
University, Egypt

ABSTRACT

Optimization becomes a very important methodology appear in scientific life. It can be applied in many different application fields, like telecommunications, data mining, design, combinatorial optimization, power systems and Electronic circuits. Development of electronic circuit is a complex process that needs some simplification that may be difficult to be done using traditional way. In This paper a hybrid rough particle swarm optimization (HRSO) algorithm is proposed for electronic circuit simplification. The (HRSO) is applied to simplify circuit by reducing the components of circuit to try to find optimal value of circuit components.

General Terms

Soft computing, data mining.

Keywords

Rough set, particle swarm optimization and electronic circuits.

1. INTRODUCTION

1.1. Reduction of data

Reduction of the knowledge necessary of inquiry into all matters of equivalence relation in space and rounded way to delete knowledge is superfluous. Dependency of knowledge presents necessary information for Knowledge reduction. The reduct is the minimum information of attributes that present the same mean of the whole set of attribute. Reduction and core are the two main tasks in basic concepts in the application of rough set theory. The knowledge reduction is the basic of knowledge that is enough to identify the concepts of this knowledge [1].

The technique of reduction is apply by Fuzzy set and rough set ,in this paper we use rough set in reduction.

1.2. Optimization processes

Optimization is defined as a math tool that makes the thing is the best. It is a process of adaptation or the characteristics of the input device, the operation of mathematical, or experiment to search for the minimum or maximum desired result. The inputs are variants, the operation of Optimization is defined the cost function, objective function, or fitness function, and the required output is the cost or fitness. If the process is the experience, then the variables are the inputs material for the experience [2].

The set of optimization techniques called Evolutionary algorithms (EAs) has been shown to be the best efficient

techniques to solve the optimization problems, so it is understanding mathematical satisfactory and accurate performance is an important challenge in the domain of evolutionary computing. There are some (EAS) have a wide range in optimization problems such as Genetic Algorithm (GA) and particle swarm optimization (PSO) [3].

Discussed the similarity and differences of three very similar evolutionary algorithms: particle swarm optimization (PSO), genetic algorithm (GA) and differential evolution (DE). In which (GA) is more appropriate in separate optimization, PSO and Differential Evolution (DE) are more natural for continuous optimization in [4].

This paper shows that hybrid PSO algorithm and GA algorithm that was proposed was giving better results in minimizing the cuts of VLSI circuit problem than used PSO algorithm separately or GA algorithm in [5].

The comparison between (PSO) and (GA) is easy to implement because they are similar in concept in which they derived from the flocking behavior of birds. They are beginning with a population of random particle. Each particle has its own position and velocity and they can be updated in iteration to the personal best position that found so far and the global best position that found so far. The advantages of PSO are that reduce the time of computation, fast convergence , it has a few parameters and PSO has a simple concept so it is easy to implement, so that in section 3 gives a brief introduction of PSO [6].

2. ROUGH SET

A rough set theory is a mathematical tool was proposed by pawlak in the early 1980's that can treat with vagueness and incomplete data. This theory can present a form for data mining to arrive the data into knowledge.

The foundations of this theory show that the concepts of learning result from examples, so the search for attributes that have certain characteristics, or decision rules depending on the data is simple and easy.

The rough set theory is important technique in data analysis and has applied applications in many fields, such as pattern recognition, machine learning, environment, engineering and expert systems. Many of applications concerned with the ideas of rough set theory as medical data analysis, image processing and voice recognition. Rough set theory can deal with data mining that has some problems such as noise, and treated values or errors due to

measurement tools inaccurate unknown that deal with this problem and treat with the incomplete and uncertain data and then gets the decision rules [7].

2.1. Rough set features

When we use Rough set theory cannot need any superfluous information, it reduces the original data, it can deal with incomplete data and express it as an efficient data analysis tools. It also reduct data that has the same knowledge as in original data and gets sets of rules from data.

Rough set theory gives powerful result when it can be used in applications alone but when it combined with other techniques such as fuzzy theory, PSO algorithm, neural networks, etc. so they get an efficient result [8].

2.2. The basic concept

The concept of rough set theory is to deal with the data and information in table that can be represent the decision table and this data represent conditional attributes and decision making attributes. Let $IS = (U, C \cup \{d\})$ be represented by a table that is called an information system, this table consists of U that is a set of finite objects that can be defined as $U = \{X_1, X_2, \dots, X_n\}$ and called the universe where every case of this object is represented by row ($U \neq \emptyset$), C is a finite set of conditions in which these conditions are represented by column ($C \neq \emptyset$), the elements of C are called conditional attribute in which $a_i \in C$ where

($1 \leq i \leq n$), a_i is called the value of attribute set and d is the decision attribute where d does not belong to C [9],[10], [11].

Representation of the data set as a table where each row represents a case, an event, a patient. Every column represented an attribute, property, variables that can be calculated for each object [1]. Suppose that $X \subseteq U$ and $P \subseteq C$, X is a subset of universe and P is a subset of the conditions, the P-lower approximation is a set of all the data that can totally belong to the principle of X that is denoted by $P_*(X)$:

$$P_*(X) = \{x \in U : \{x\}_p \subseteq X\}$$

The P-upper approximation is the set of all data that can possible belong to the principle of X that is denoted by $P^*(X)$:

$$P^*(X) = \{x \in U : \{x\}_p \cap X \neq \emptyset\}$$

The P-boundary region of X is: $PN(X) = P^*(X) - P_*(X)$.

X is said to be roughly if and only if $P^*(X) \neq P_*(X)$ and boundary region $\neq \emptyset$, X is said to be definable

If and only if $P^*(X) = P_*(X)$ and boundary region = \emptyset [11], [12].

3. PARTICLE SWARM OPTIMIZATION

Particle swarm optimization (PSO) is evolutionary optimization algorithm. This approach can depend on the movement and intelligence of swarm, which was first developed by James Kennedy and Russell c. Eberhart in the mid-1990s. the PSO technique is resulting from the social behavior of birds, fishes and ants. This algorithm has some features, it is simple in concept so it is easy to learn, quick in implementation and can be coded in a few

lines on computer, this approach gives the optimal solution in a few time. Each particle in PSO has two storage positions: first the local best position that is the best solution known as (pbest), second the global best position that is a set of best solutions known as (gbest) [13].

PSO starts with a population of N particles. Each particle X_i has its own position then it transmits to a new position in the space with velocity V_i .

In which $X_i^t = \{X_{i1}^t, X_{i2}^t, \dots, X_{iL}^t\}$ represents the position i of component L in iteration t , and the $V_i^t = \{V_{i1}^t, V_{i2}^t, \dots, V_{iL}^t\}$ represents the velocity of particle i of component L in iteration t . Suppose that $pbest_i^t$ represents the local best position that is arrived by the particle i so far. Where $gbest_i^t$ represents the global best position arrived by any of the set of particles that are neighborhood to local best position so far.

Then the velocity of each iteration can be calculated by this equation:

$$V_{ik}^{t+1} = [wV_{ik}^t] + [w1c1][(pbest_{ik}^t) - (X_{ik}^t)] + [w2c2][(gbest_{ik}^t) - (X_{ik}^t)] \quad \text{Eq[1]}$$

In which, t : is the number of iteration that arrived to it so far, i : is number of particle that arrived to it where

$i = \{1, 2, \dots, N\}$, K : is the component index where

$K = \{1, 2, \dots, L\}$, $w1$ and $w2$ are two random numbers their values are between the interval $[0: 1]$. In each iteration $c1$: controls the impact of local best position and $c2$: control the impact of global best position. w : is a positive parameter called inertia weight, which can control on the velocity of the previous iteration to avoid exploration in population of swarm.

After calculating the velocity of particle i then the particle can move to a new position with this equation:

$$X_{ik}^{t+1} = X_{ik}^t + V_{ik}^{t+1} \quad \text{Eq [2].}$$

In which the new position of particle i in iteration

($t + 1$) Can be calculated by adding a velocity of particle i in iteration ($t + 1$) to the current position [14],[15].

Fitness function:

The fitness function is used to calculate the quality of the position of each particle.

The fitness value of this circuit is evaluated by this equation:

$$v_{out} = (v_{in1} - v_{in2}) \left[\left(1 + \frac{R2+R3}{R1} \right) \left(\frac{R6+R7}{R4+R5} \right) \right] \quad \text{Eq[3].}$$

4. RELATED WORK

4.1. Rough set theory

In [8] suggested rough set theory combining with a high-risk customer management model to solve the problem of traditional risk management that cannot deal with historical data efficiently. This paper gives a good experience to show how the development and application of the proposed model and then get a decision rules without any Supplementary information.

In [16] presented new generalized concepts of lower and upper approximations of rough sets by using two

topological structures, they discussed the concept of the topological rough set and some of their important properties and then they applied the reduction on medical data.

In [17] suggested rough set theory to analyses the data samples error of high voltage circuit-breaker and find diagnosis rules of it.

4.2. Particle swarm optimization (PSO)

In [18] presented an approach based on simulation to improve the performance of a series of folded OTA by using particle swarm optimization algorithm. There exists in OTA design a relationship between Gain, BW, PM, CMRR and PSRR which force the designer to investigate such a complex work a long time to improve the circuit. In this paper has been to investigate the use of the PSO algorithm in the optimization of automated FCOTA.

In [19] a suggested (PSO) algorithm that is suitable for digital combinational logic circuit to reduce the number of their gates. They used MATLAB, to code the circuits into particles by using this algorithm. PSO applied to optimize a full-adder circuit; the result shows that a full-adder circuit has been minimized from six gates that obtained from human-designer method to five gates.

In [20] proposed the Particle Swarm Optimization (PSO) algorithm to solve problem of reverse gravity, the presented solution to the inverse problem is determining the form of error which gravity anomaly is known.

In [14] an applied (PSO) algorithm to specify the optimal type and location of new wells that is a key element in the performance of Development of oil and gas fields.

In [15] suggested modified PSO algorithm by using the idea of simulated annealing algorithm. That makes the optimum particle of all time repetitions are constantly evolving, and the worst set of particles with a new value to increase its disturbance. The result shows that the modified PSO algorithm has better efficiency in convergence than the original PSO.

In [21] presented a new PSO approach to control the search difficulties. This approach discussed two search strategies for multimodal functions. One of the cooperative Strategy, which controls the area and the search is integrated partial and complete the remoteness of the PSO. The other is to overcome the speed and Particles in the way of adaptation, according to the movement of it in Space. The experiments were carried out to evaluate the suggested approach, and the results show that this approach can lead better than other variables PSO popular approach.

4.3. Hybrid rough set with (PSO)

In [22] suggested a rough set which is the best feature selection method to reduce the database that has huge number of features, but this method cannot give the optimal solution so this paper hybrid Rough set theory with Bee Colony Optimization (BCO) in order to find optimal solution with no random parameter assumption.

In [23] presented algorithm based on hybrid of Particle Swarm Optimization (PSO) and Ant Colony Optimization (ACO) algorithms in order to promote improvement Performance. This approach is used in the reduction computation in rough set to find the optimal fitness value.

In [24] introduced overview of many PSO approaches that is used for partitioning an image into different meaningful segments, PSO is a good performance technique in this domain but when combined with other techniques such as rough set, fuzzy sets, neural network and so on, they are gives more efficient solution.

5. PROBLEM FORMULATION

Our problem focuses on the complexity of electronic circuit design that can need to be in simple design and then try to find the optimal solution. The simplification and optimization processes are more difficult to be done by traditional ways, so we can use HRSO algorithm for reduction and optimization. We can reduce the components of circuit without effect on the target and then applied the optimization, so our goal for reduction is to reduce the components that cannot effect on the output so it makes the circuit simpler therefore reduce the cost. And the aim of optimization is to find proper values of the remaining components of this circuit in a few iterations.

Fig.1 [25] shows an instrumentation amplifier diagram in which the circuit has two parts: the first one is U5,U6. It is called buffer amplifier the second part U1 is called differential amplifier in which this part is the basic of amplifier.

6. (HRSO) ALGORITHM

Input: number of conditions C_k , number of objects X and decision d .

Output: reduct set of conditions.

1- remove $C_k \forall K = 1, 2, \dots, 7$.

//computing upper and lower approximation set.

2-set $P^*(X) = \phi, P_*(X) = \phi$.

3-If $X_i = X_j, d_i = d_j$,

then $X_i = \{ \}$.

4-elseif $X_i = X_j, d_i \neq d_j$, Then $X_i, X_j \subset P^*(X)$.

5-else $X_i \neq X_j$, Then $X_i \subset P^*(X), P_*(X)$.

6-end if.

7-If $P^*(X) = P_*(X)$, then C_k is consistent.

8-else C_k is inconsistent.

9- End if.

//algorithm of PSO

10- Initialize randomly parameters w_1, w_2, c_1, c_2 .

11- Initialize the number of particle N and number of maximum iterations.

12-Initialize randomly position vector X_{ik}^t and velocity vector V_{ik}^t

13- Calculate fitness value of each particle i from fitness function by Eq[3].

14- Initialize $pbest_{ik}^t$ in first iteration and determine $gbest_{ik}^t$.

15-update the velocity by Eq[1] in each iteration.

16- Then update the position by Eq[2] in each iteration.

17- \forall particle i , if $f[X_{ik}^t] > f[pbest_{ik}^t]$, then

Update $pbest_{ik}^t = X_{ik}^t$.

18- \forall particle i , if $f[X_{ik}^t] > f[gbest_{ik}^t]$, then

Update $gbest_{ik}^t = X_{ik}^t$.

19- Repeat iteration until result fixed or the maximum iterations exceeds if not go to step 6. [9]

7. RESULT

After applying HRSO algorithm, rough set theory removed three resistant (R1, R4, R7)but with the same output so this resistant cannot effect on our output of the circuit. Then the instrumentation amplifier becomes more simplification show in Fig2. particle swarm optimization (PSO) converges much faster and achieves better fitness value that can reach to the optimal solution after 11 iterations compared this with manual way that can reach after 117 iterations, So The fitness value versus the number of iterations in the optimization of instrumentation amplifier is shown in Fig.3.

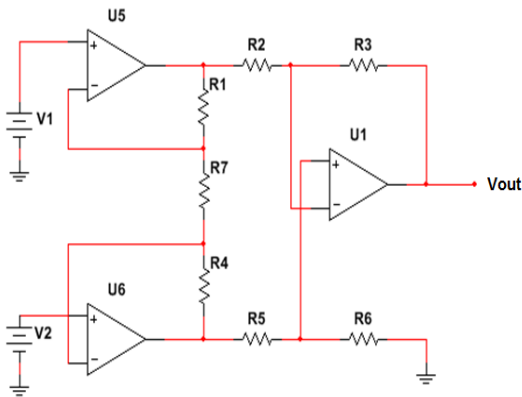


Fig.1 Typical instrumentation amplifier schematic

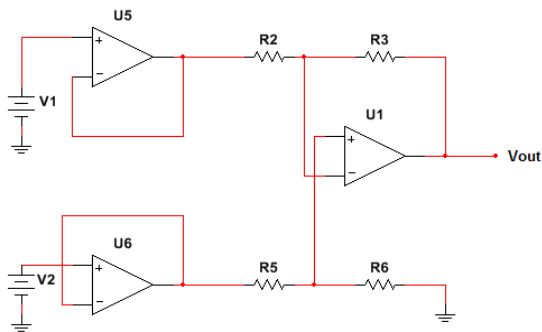


Fig.2: after applying rough set.

8. CONCLUSION

We applied HRSO algorithm on instrumentation amplifier, this approach was applied for the reduction and optimization. In which rough set was applied three resistant (R1, R4, R7) were be removed from the components of the circuit without any effect on the output of our circuit so rough set is efficient and more powerful in reduction components of the circuit therefore make it in simple design and reduce the cost. PSO was applied so it can be optimized the values of the remaining components and reached to the optimal solution in a few iteration so PSO has two advantages it easy in implementation and fast

convergence, so HRSO algorithm has a high performance and good effective in the reduction and optimization.

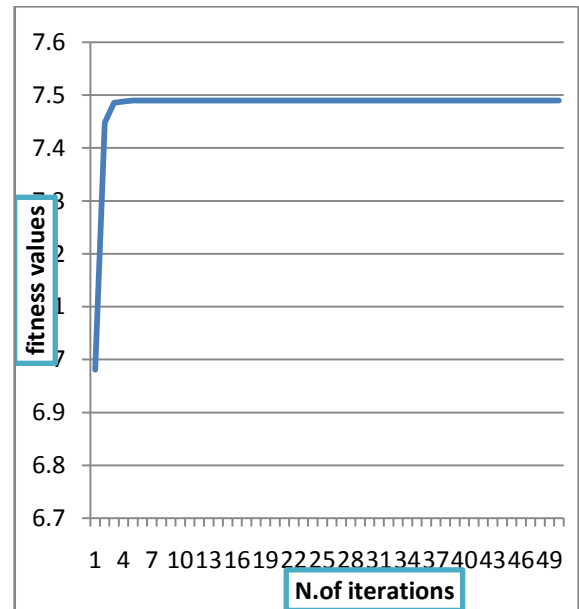


Fig.3. Fitness values versus the number of iterations in the optimization of circuit

9. ACKNOWLEDGMENTS

We sincerely thank the reviewers for their valuable comments that improve the presentation of the paper.

10. REFERENCES

- [1] Shi;Zh.,”advanced artificial intelligence”, world scientific,Vol.1, pp.393,2011.
- [2] Haupt;R.,and Haupt;S.,” Practical Genetic Algorithms”, second edthedn.,John Wiley & Sons, New Jersey, pp.1-2,2004.
- [3] Doerr;B.,” Evolutionary Algorithms and Dynamic Programming”,pp.2-34,2013.
- [4] Kachitvichyanukul; V.,” Comparison of Three Evolutionary Algorithms: GA, PSO, and DE”, Industrial Engineering& Management Systems, Vol 11, No 3, pp.215-223,2012.
- [5] Singh;R.,et.al,“Hybrid Optimization Technique for Circuit Partitioning Using PSO and Genetic Algorithm”, IJETEE,Vol. 4, Issue. 2, 2013.
- [6] Behera;H.,” Segmentation and Classification using Heuristic HRSPSO”, J. (IJSCE),Vo11 ,Issue 3 ,pp.66-69,2011.
- [7] Venugopal;K.,et.al., “Soft Computing for Data Mining Application”, SCI 190,p 16.,2009.
- [8] Zhiyao;L.,et.,al.“ High Risk Management Model For The Power Enterprise Based on Rough Set Theory”, Systems Engineering Procedia,Vol 3 ,pp.63 – 68,2012.
- [9] Chu.Y., et.al,“Study on Fault Diagnosis of Circuit-breaker Based on Rough-Set Theory”,TELKOMNIKA, Vol.11, No.1, pp. 296-301,2013.

- [10] Ahmed;T.,et.al.,“Data Missing Solution Using Rough Set Theory and Swarm Intelligence”, JIJACSIT, Vol. 2, No. 3, Page: 1-16, ISSN :2296-1739,2013.
- [11] Renu; V., and M.L. Garg,“ computing the significance of an independent variable using rough set theory and neural network”,IJREAS ,Volume 3, Issue 3, ISSN: 2249-3905,pp. 122-136,2013.
- [12] Yu;J.,et.al.,”AndrzejSkowron Rough Set and Knowledge Technology”, 5th International Conference, pp.135 ,2010 .
- [13] Mehnen;J., et.al.,”Applications of Soft Computing:from theory to praxis”, Intelligent and Soft Computing 58, Springer, pp. 502-503,2009.
- [14] Onwunalu;J., and Durlofsky;L,“Application of a particle swarm optimization algorithm for determining optimum well location and type”,J. ComputGeosci ,pp.183–198, 2010.
- [15] Mu;A.,et.al.,” A Modified Particle Swarm Optimization Algorithm”,natural science,Vol.1, No.2,pp. 151-155 ,2009.
- [16] Salama ;A., “Bitopological rough approximations with medical applications”, J.King Saud University (Science) 22,pp. 177–183,2010.
- [17] Yan;C.,et.al., “Study on Fault Diagnosis of Circuit-breaker Based on Rough-Set Theory”,TELKOMNIKA, Vol.11, No.1, pp. 296-301,2013.
- [18] Nakhaei;R., and Zahaby;M.,” Performance Optimization of Folded Cascode OTA Using an Evolutionary Algorithm”, J. Computer Science and Software Engineering , Vol 3, Issue 6,pp.60-64,2013.
- [19] Orgi;J.,et.al.,“ Optimising digital combinational circuit usingparticle swarmoptimisation technique”,J.phys.Educ.,Vol.6,No.1,pp.27-77,2012.
- [20] Touthmalani ;R,“Gravity inversion of a fault by Particle swarm optimization (PSO)”, Touthmalani Springer Plus ,pp.1-7, 2013.
- [21] Yang Wang;R.,et.al., “A New Cooperative PSO Approach for the Optimization of Multimodal Functions”, WCE, Vol2,pp.1-6, 2012.
- [22]Suguna;N., and Thanushkodi;K.,” A Novel Rough Set Reduct Algorithm for Medical Domain Based on Bee Colony Optimization”, J. Computing ,Vol. 2 ,pp. 49-54 ,2010.
- [23] Pratiwi.L, et.al., “Improving Ant Swarm Optimization with Embedded Vaccination for Optimum Reducts Generation”, IEEE,Vol2 ,pp.448-453,2011.
- [24] Kaur;A.,and Singh;M.,“ An Overview of PSO- Based Approaches in Image Segmentation”, J. of Engineering and Technology, Vol 2,pp.1349-1357,2012.
- [25] Bali,"Linear Integrated Circuits"p.p.224,2009.