

FUZZY GENE OPTIMIZED REWEIGHT BOOSTING CLASSIFICATION FOR ENERGY EFFICIENT DATA GATHERING IN WSN

J.Srimathi¹ and V.Valli Mayil²

¹Ph.d, Research Scholar, Bharathiar University, India

²Head & Associate Professor, Dept. of Computer Science & Applications, Periyar Maniammai University, Thanjavur, India

ABSTRACT

The energy is a major resource to obtain efficient data gathering and increasing network lifetime (NL). The various techniques are introduced for data aggregation, but energy optimized sensor node (SN) selection was not carried out to further enhance NL. In order to improve the energy efficient data gathering in WSN, a Fuzzy Gene Energy Optimized Reweight Boosting Classification (FGEORBC) Technique is introduced with lesser time consumption. In FGEORBC technique, the Residual Energy (RE) of SN in the WSN is computed. After calculating SN residual energy, fuzzy logic is applied to determine higher energy nodes and lower energy nodes using threshold value. For finding the optimal higher energy SNs, then Ranked Gaussian gene optimization technique is applied. If the node satisfies the fitness criterion, then the node is selected as an optimal higher energy SN. Otherwise, the rank selection, ring crossover, and Gaussian mutation process are carried out until the condition gets satisfied. After that, the sink node collects the data packets (DP) from the optimal higher energy SNs. In the sink node, Reweight Boosting Classification is carried out to classify the sensed DP and it sends to the base station (BS) for further processing. Simulation of FGEORBC technique is carried out using different parameters such as energy consumption (EC), NL, data gathering time and classification accuracy (CA) with respect to a number of SN and a number of DP. The results observed that FGEORBC technique improves the data gathering and NL with minimum time as well as EC than the state-of-the-art methods.

KEYWORDS

WSN, data gathering, residual energy, fuzzy logic, Ranked Gaussian gene optimization, data classification, Reweight Boosting Classification

1. INTRODUCTION

Wireless Sensor Network (WSN) comprises a group of sensor nodes (SN) for sensing and collecting the physical circumstances. The distributed SN has the ability to gather the sensed data and transmit to the sink node. WSNs are generally self-organized wireless ad-hoc networks which include a large number of resource-limited SNs. The major tasks of distributed SN are, gathering efficient data and to send a remote BS. SN is dispersed in an environment and uses batteries as an energy resource. In order to carry out several tasks, the mobile nodes need additional energy. The higher energy utilization of the SN minimizes the NL. Therefore, the NL becomes an imperative metric for sensor network to improve the data collection performance. Energy efficiency is a vital issue in WSN as sensor nodes are battery-powered. The data aggregation is a significant process to save energy by avoiding redundant data broadcasting in WSNs. To increase energy efficiency and make longer the network lifetime of the sensor networks, a lot of protocols is designed in conventional works. However, energy efficient data gathering is not obtained. Therefore, FGEORBC Technique is proposed in this research work.

In [1], a Distributed Data Gathering Approach (DDGA) was introduced to minimize the data collection problem with a mobile sink. But, the sink node failed to classify the collected DP for further processing in WSN. A mobile agent based compressive data gathering (MA-Greedy) approach was introduced in [2] for minimizing the EC using Coefficient of Variation. The MA-Greedy approach minimizes the time complexity but it failed to determine the high and low energy SN to extend the NL.

The Energy-Efficient Data Gathering approach was introduced in [3]. The approach measures the relationship among sensing information of the SNs, it failed to use efficient classification method to classify collected data. An energy efficient structure-free data aggregation and delivery (ESDAD) protocol was introduced in [4] for gathering the data with minimum delay. The protocol does not obtain reliability in the data aggregation.

Distributed Trajectory Design was presented in [5] for collecting the data with Mobile Sink for improving the NL and minimizing the EC. But the sink node failed to classify the data for further processing in WSN. A modified Expectation-Maximization (EM) clustering technique was introduced in [6] for collecting the data with minimum EC. The clustering technique failed to handle more SN with high gathering delay.

An auction-based scheme was introduced in [7] for improving data aggregation with minimum energy utilization and hence enhanced the NL of the SNs. Though the method minimizes data loss, the data classification was not performed.

A Residual Energy Aware Mobile Data Collection method was developed in [8] to enhance NL with minimum delay. The method consumes more time for gathering the data from high energy SN in WSN. Markov Decision Process was introduced in [9] for energy-efficient data collection but it takes a longer delay.

In [10], a Cell-based Path Scheduling (CPS) approach was developed to collect the snapshot data. But, the algorithm failed to compute the RE of the SN before the data collection.

The major problems are identified from the above-said literature such as lack of accurate data gathering, high time complexity, and failure for data classification, high EC, and failure to enhance the NL and so on. In order to overcome such kind of issues, an efficient FGEORBC technique is developed in WSN.

The main contributions of FGEORBC technique are summarized as follows,

- FGEORBC technique improves energy-efficient data gathering with less time consumption. FGEORBC technique uses fuzzy logic to find the SN energy level. The nodes with greater RE are chosen to minimize the EC.
- To improve the NL, FGEORBC technique uses Ranked Gaussian gene optimization for selecting the high energy SN based on the fitness measure. Depends on fitness value, an optimal high energy SN is selected. If the node does not satisfy the fitness criteria, the rank selection is applied to choose the current best individuals (i.e. nodes). Then the ring crossover is performed to swap the input string for generating the offspring. The Gaussian bit string mutation is performed to invert the bit strings. Finally, the fitness condition of the newly generated individual is verified to select the high energy SNs.
- The sink node is a data aggregator which collects the data from higher energy SN with minimum time. Then, FGEORBC technique classifies the data using reweight boosting

technique. Initially, a decision stump is constructed based on the correlation between the DP. Decision stump is a one level decision tree which classifies the DP using Bivariate correlation. The weak learner's results are combined to provide strong classifier with minimum error. This aids to improve the CA. Finally, the classified DP sent to the BS.

The rest of the paper is organized into five different sections. In section 2, the review of the related works is presented along with their ability and limitations are discussed. Section 3 provides an overview of the proposed FGEORBC technique workflow with neat diagram. In section 4, the simulation settings are performed with certain simulation parameters. The simulation results of FGEORBC technique compared to other conventional schemes are presented in Section 5. Finally, the conclusion of the paper is presented in Section 6.

2. RELATED WORKS

High-throughput opportunistic data collection approach was presented in [11]. The approach performs efficient data collection but it failed to categorize the DP at the sink node. An Immune-Genetic heuristic algorithm was designed in [12] for improving the accuracy of the aggregated data received by the sink node. But the algorithm takes more time for data aggregation.

A new Data Routing for In-Network Aggregation method was developed in [13] for improving the performance of data gathering and communication. The method failed to consider the spatial and sequential correlation of the aggregated data for classification at the sink node.

Multi-Strip Data Gathering (MSDG) approach was introduced in [14] for improving the performance of data collection and NL by utilizing the minimum energy. The approach failed to use an adaptive distributed technique for minimizing the complexity in data collection. A Mixed-integer linear programming (MIP) model was designed in [15] for increasing the optimal data-gathering and NL. The model did not exploit any optimization technique to discover higher and lesser energy SNs.

A mobile agent protocol based Energy-aware data collection method was introduced in [16] to improve the data collection with minimum energy as well as time. But the performance the NL remained unaddressed.

Mobile cluster-head data collection model was presented in [17] for improving the packet delivery and minimizing the end to end delay. But this model failed to balance the non-uniform flow of data. Maximum Lifetime Data Aggregation Tree Scheduling approach was introduced in [18] to maximize the NL. The approach does not perform accurate data aggregation in dynamic networks.

A dynamic mobile agent based data aggregation method was introduced in [19] for enhancing the NL by lessening the EC. However, this method failed to support the various cooperative agents for enhancing the aggregation precision and reducing the delay. An energy efficient hybrid approach called cluster and tree-based data gathering was presented in [20] to improve the NL. However, the time complexity was more.

An energy efficient data aggregation protocol was intended for decreasing the transmission path and direct communication in WSN. But, the latency of the network was not solved. Secure Clustering Protocol was presented in the network lifetime.

The impact of duty cycle on energy utilization in WSN was analyzed. This model was introduced to discover communication delay and energy usage in WSN. But, network lifetime was lower. An enhanced Stable Election Protocol was presented in Insensitive Fuzzy C-Means to increase the energy efficiency of the network. Balanced Energy Efficient Network Integrated Super Heterogeneous (BEENISH) was designed.

To overcome the above said existing issues, FGEORBC technique is designed and presented in the next section with a neat diagram.

3. A FUZZY GENE ENERGY OPTIMIZED REWEIGHT BOOSTING CLASSIFICATION TECHNIQUE FOR DATA GATHERING IN WSN

Data aggregation is the process of collecting the data from various SN and transfers the information to the BS in an energy efficient manner with lower data latency. Energy efficient data aggregation plays a vital role in enhancing the NL. This also improves the functionality of the SN in the network. In energy limited sensor networks, it is inefficient for sensors to transmit the data directly to sink at a longer period of time. In such scenarios, SN transmits the data to a local high energy aggregator which aggregates data from all the low energy SNs. Therefore, the major concern in the design of a WSN application is to improve the data gathering efficiency of the SN with lesser energy utilization. Based on this motivation, FGEORBC technique is introduced.

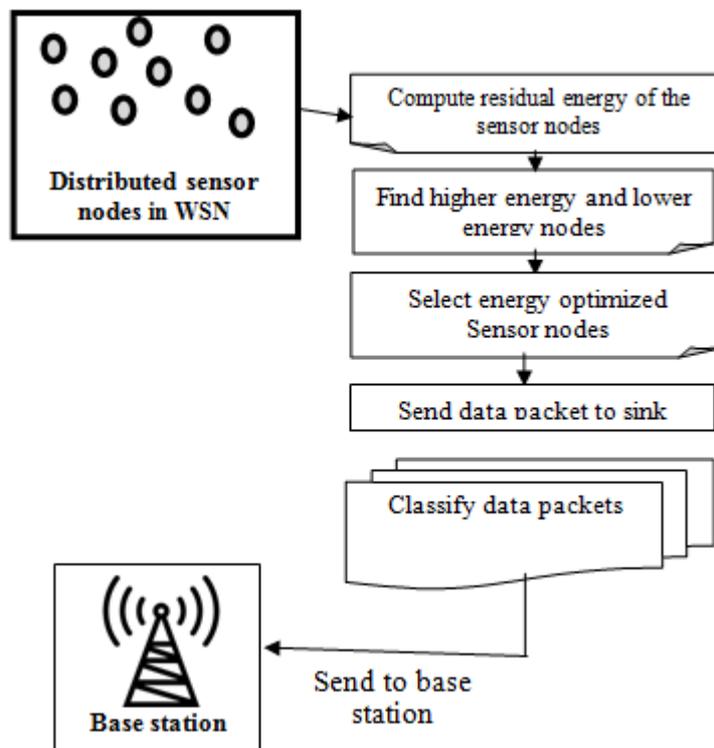


Figure 1. Flow process of the FGEORBC technique

Let us consider the WSN designed in the graphical model $g(v, e)$ where v denotes a number of SNs. All the SN is allowed to monitor environmental conditions and continue forward to the base station (b_s). In the graph, the connections between the SN are denoted as ' e '. The sink node collects the data dp_1, dp_2, \dots, dp_n and sends to the BS for further processing. In the data gathering

process, the SN energy is considered for increasing the network functionality. The proposed FGEORBC technique is designed with the above system model. The FGEORBC includes three major processes to improve the efficiency of data gathering in WSN with less EC. These three processes are illustrated in the architecture diagram 1.

Figure 1 illustrates the flow process of FGEORBC technique to obtain energy-efficient data gathering in WSN. Let us consider a WSN that includes many SN within a limited transmission range. At first, entire SN has an equivalent energy level. Then the SN performs two processes namely sensing and collecting the data. After the sensing, the energy of the sensor nodes gets minimized. In this case, the RE of each SN is computed. The FGEORBC technique finds the high energy nodes by applying the fuzzy logic. After finding the higher energy nodes, the low energy node finds the nearest higher energy node using Ranked Gaussian gene optimization technique to transfer the collected data. After that, the sink node gathers sensed data from the higher energy nodes and classifies the data using reweight boosting classification technique. Finally, the sink sends the classified data to the base station. The detailed explanations of these three processes are explained in following subsections.

3.1 Fuzzy logic based nodes energy evaluation

The first process in the FGEORBC technique is to determine the high energy and low energy nodes among the distributed SN in WSN using fuzzy logic. A fuzzy logic uses a rule for connecting the inputs with the outputs. By the established rules, the mobile nodes and their energy level are identified. The rules are formulated as IF (condition) and THEN (conclusion). The condition part checks the energy of the mobile nodes with the threshold energy level. The conclusion part provides the desired outputs. Based on this rule, the high and low energy SN are identified. The energy of the SN is measured in joule (J). Then, the SN monitors the environmental conditions for collecting the data. RE of SN is determined after the sensing process.

$$r_e = t_e - c_e \quad (1)$$

From (1), r_e signify residual energy, t_e signify total energy and c_e signify a consumed energy of SNs. Hence, the remaining energy is computed as a difference between the total energy and the consumed energy of the SN. After calculating the SN's residual energy, fuzzy logic is applied to find the higher energy nodes and lower energy nodes with the threshold energy level. If the node has high RE than the threshold, then the particular node is considered as higher energy node. Whereas The RE is lower than threshold energy level (δ), the SN is considered as lower energy nodes. Then, the FGEORBC technique detects high energy and low energy nodes. The algorithmic process of fuzzy logic based node selection is described as follows,

```

Input: Number of sensor nodes  $sn_1, sn_2, sn_3, \dots, sn_n$ 
Output: Find high energy and low energy sensor nodes
Begin
  1. for each sensor node  $sn_i$ 
  2.   Measure similar energy  $e_g$ 
  3.   Compute residual energy  $r_e$ 
  4.   if ( $r_e \geq \delta$ ) then
  5.      $sn_i$  is said to be higher energy nodes
  6.   else
  7.      $sn_i$  is said to be lower energy nodes
  8.   endif
  9. end for
end

```

Algorithm 1. Fuzzy logic based energy efficient node selection

Algorithm 1 describes the energy efficient node selection using fuzzy logic approach. Initially, the entire SN in the network contains equivalent energy. Once sensing the data, remaining energy of SN is computed, RE of SN is greater than the threshold, and then the SN is considered as high energy node. Otherwise, the SN is said to be lower energy SNs. By this way, FGEORBC technique finds the energy efficient SN for data gathering as well as improving the NL.

The high energy SN extend the network life but the low energy SN does not maintain the data at a longer period of time. The lower energy SN needs additional energy for transmitting the sensed DP to the sink node. To resolve the issues, the lower energy SN finds the nearest energy efficient nodes using gene optimization algorithm. The optimization process is described in the next subsection.

Table 1. Notations used in the FGEORBC technique

Symbol	Definition
sn_i	Sensor Node
v	a number of SNs
e	connections between SN
b_s	base station
r_e	residual energy
t_e	total energy
c_e	the consumed energy of SN
δ	threshold energy level
f	fitness function
D	distances
(x_1, y_1) and (x_2, y_2)	current coordinate of sn_1 and sn_2
p_s	selection probability
$r(b)$	rank of the individual
$\sum_{i=0}^n r_i$	the sum of rank values of the entire population
S	newly generated offspring value
v	variance
'r'	Gaussian normally distributed a random number
x	input DP
y	classification output
$w_i(x)$	number of weak learners output
γ	correlation coefficient
n	total number of data packets
$H(x)$	the output of the weak learner
α	the weight assigned to weak learners
e	error
y_i	actual output
Y	the output of the strong classification result
α^w	Reweight of the weak learner ' $w_i(x)$ '
$\arg \min e$	minimum error

3.2 Ranked Gaussian Gene Optimization based energy efficient node selection

The second process of the FGEORBC technique is to select the energy optimized SN for sending the DP. The ranked Gaussian Gene optimization algorithm is used to find the nearest energy

efficient node. It is a Meta heuristic technique provides a better solution for the optimization. Initially, the sensor nodes $sn_1, sn_2, sn_3, \dots, sn_n$ populations are generated in search space. In each generation, the fitness of every SN is computed. The fitness is the value of the objective function in the optimization problem being solved. Based on the fitness calculation of the SNs, the more fit individuals are stochastically selected from the current population. The fitness of the node is computed as follows,

$$f = high r_e + \min D \quad (2)$$

From (2), f denotes a fitness function which is computed based on the high residual energy (r_e) and minimum distances (D). The lower energy SN finds the high energy nearest SN with the minimum distance. The distance between the nodes is computed using Euclidean distance. The current coordinate for sn_1 is (x_1, y_1) and for sn_2 is (x_2, y_2) . Distance among two SN is computed to discover the nearest node. The distance is measured as follows

$$D = \sqrt{\sum_{i=1}^n (y_i - x_i)^2} \quad (3)$$

From (3), D denotes a distance between the SNs sn_1 and sn_2 . Therefore, the node with minimum distance is selected as the nearest energy efficient node and the sink node gathers the data from this higher energy node for improving the NL. If the current solution is not satisfies the fitness criterion, the genetic parameters such as the rank selection, ring crossover and Gaussian mutation process is carried out until the condition gets satisfied.

3.2.1 Rank Selection

Selection is the first genetic operation and it helps to select the best mobile nodes in the population for creating the offsprings the next generation. In the selection process, two parent chromosomes are chosen from the initial population using rank approach. In this selection process, the chromosomes are arranged based on the fitness value. After that, the individuals are ranked. The selection probability is obtained by the ratio of a rank of the individual chromosomes and the sum of rank values of entire populations. Then the probability is computed as follows,

$$p_s = \frac{r(b)}{\sum_{i=0}^n r_i} \quad (4)$$

From (4), p_s represents a selection probability, $r(b)$ denotes a rank of the individual. $\sum_{i=0}^n r_i$ denotes a sum of rank values of the entire population. The individuals are represented by an array of bits, where each bit value is 0 or 1. Based on the parent chromosomes are chosen along with their selection probability for the next crossover process.

3.2.2. Ring crossover

After the selection process, the genetic information is interchanged between two individuals. The information swapping in genetics is called as crossover and newly generated chromosomes are called offspring. The FGEORBC technique uses the ring crossover for generating the offspring from the parent chromosome resulting in generates a second generation population of the solution. This ring crossover performs four steps for creating the offspring. In the first step, the two parent chromosomes are defined which is chosen by the rank selection.

$$A = 1011100$$

$$B = 111010$$

Two parent chromosomes *A* and *B* are selected. In the second step, the parent chromosomes are arranged in a ring form. In the third step, cutting point is randomly selected for swapping the two parent chromosomes. Final step, new offspring's is generated by swapping the chromosomes using cut point.

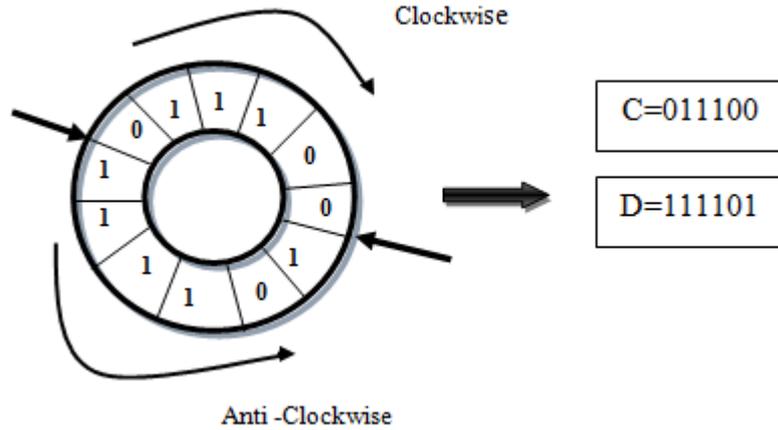


Figure 2. Offspring generation using ring crossover

Figure 2 illustrates an offspring generation from the recombination of the two parent chromosomes. After the crossover, the generated offspring's length is equal to the parent's chromosome. There are two offspring's are generated in the clockwise and anticlockwise direction as shown in figure 2.

3.2.3 Gaussian bit string mutation

After generating the offspring's, each individual's is modified and possibly randomly mutated to form a new generation. In the mutation process, the Gaussian Mutation is used to randomly interchanging the bits of the chromosome. It also helps to maintain genetic diversity from one generation of a chromosomes population to the next generation. By applying Gaussian mutation, the formula is expressed as follows,

$$S = \frac{1}{\sqrt{2\pi\delta^2}} * e^{-\frac{r^2}{2v^2}} \tag{5}$$

From (5), *S* denotes a newly generated offspring value after inverting the bit. Here, 'v' denotes a variance, 'r' denotes a Gaussian normally distributed random number with 0 mean and variance 1. Based on the Gaussian mutation operator the bits are interchanged as follows,



Figure 3. Gaussian bit string mutation

Figure 3 shows the Gaussian bit string mutation and it generates the new offspring to select optimal energy efficient SNs. After that, the before selected individual is replaced with a new individual. Then the fitness criterion is verified for finding the nearest energy efficient nodes. These processes are continued for all the possible nearest SN in a population and discover the most suitable energy efficient nodes. The flow process of the ranked Gaussian gene optimization is shown in figure 4 to choose nearest energy efficient SNs.

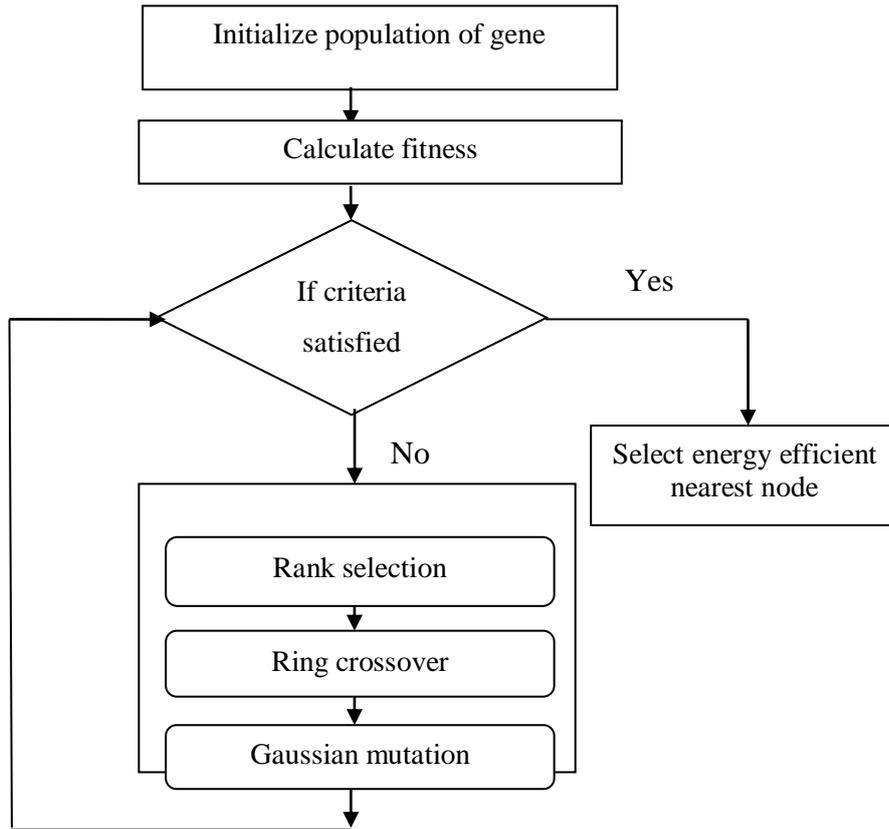


Figure 4. Flow Process of Ranked Gaussian gene optimization

Input: Number of sensor nodes $sn_1, sn_2, sn_3, \dots, sn_n$, data packets $dp_1, dp_2, dp_3, \dots, dp_n$

Output: Energy efficient nearest SN selection

Begin

1. **Initialize energy efficient** $sn_1, sn_2, sn_3, \dots, sn_n$
2. **for each** sn_1
3. **Compute the fitness** f
4. **if** (fitness criteria satisfied)**then**
5. Selects an optimal energy efficient nodes
6. **else**
7. Select the individual from the population with a probability p_s
8. Generate new offspring using ring operator
9. Perform Gaussian mutation for inverting the bit
10. Replace old individual into a new one
11. **endif**
12. Go to step 3
13. Terminate until the specified condition is satisfied
14. **end for**

end

Algorithm 2. Ranked Gaussian gene optimization

Figure 2 describes the algorithmic process of ranked Gaussian gene optimization to select the energy efficient nearest SNs. The population of the nodes is initialized randomly. Then, the fitness is computed for each individual. If the nodes satisfy the fitness criteria, then the SN is selected as energy efficient nearest node. Otherwise, the rank selection, ring crossover, and Gaussian mutation are performed. In rank selection, the parent chromosomes are selected based on the current fitness value. After that, the swapping of the two parent chromosomes is carried out to generate the offspring. Finally, the Gaussian mutation is applied to inverting the bit randomly and again the fitness criteria are verified. This process is continued until a lower energy SN discovers the nearest energy efficient node. After finding the energy efficient nodes, the low energy node sends their data to higher energy nodes. After that, higher energy SN sends the DP to the sink node.

3.3 Reweight Boosting technique for data classification

After transmitting the data, the sink node collects all the DP from the high energy SN with minimum time and it performs data classification using Reweight Boosting Classification technique. Boosting is a machine learning ensemble classification algorithm that changes the performance of weak learners into strong ones. A weak learner is a base classifier that failed to provide an accurate classification. In contrast, a boosting classifier is a strong classifier that provides accurate classification results by combing all the weak learners. As a result, FGEORBC technique uses the Reweight Boosting algorithm to improve the CA and minimizes the incorrect classification. The boosting algorithm exploits a decision stump as a weak learner to classify the data into different classes. The flow process of reweight Boosting Classifier is illustrated in figure 5.

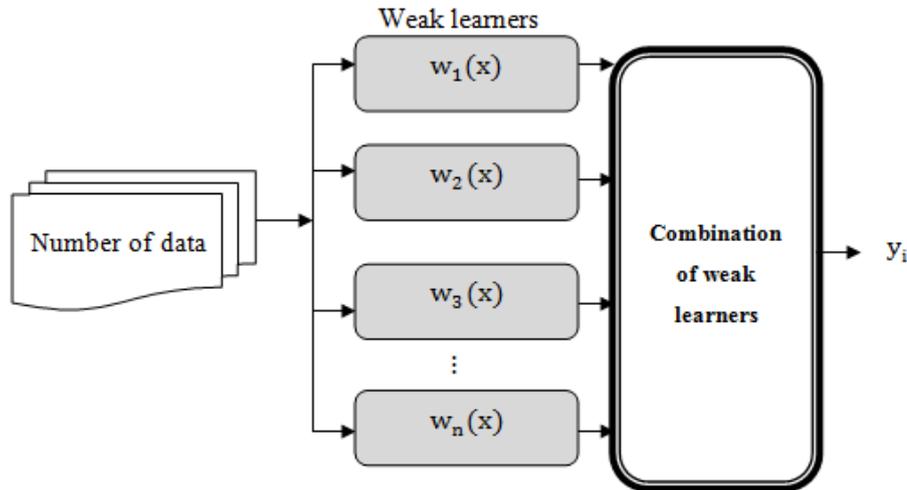


Figure 5. Structure of reweight boost classification

Figure 5 shows the structure of the reweight boost ensemble classifier to classify the data into different classes for further processing. The sink node performs the data classification and it sends to base station. Let us consider the training data $\{(x_1, y_1), (x_2, y_2), \dots (x_n, y_n)\}$ where x denotes an input DP and ‘ y ’ represents a classification output. The set of weak learners $\{w_1(x), w_2(x), w_3(x) \dots w_n(x)\}$ are constructed to train the input training data and combined into a strong one. A decision stump is a decision tree where the DP is classified based on the correlation. Correlation is a statistical method used to show how strongly the variables are related

to each other. Therefore, a decision stump performs classification based on the correlation measure. The bivariate correlation is used to determine the linear relationship between two aggregated DP. Let us consider the number of data packets $dp_1, dp_2, dp_3, \dots, dp_n$ collected by the sink node from the high energy SNs. Then the sink node measures the correlation between the DP is measured as follows,

$$\gamma(dp_1, dp_2) = \frac{n \cdot \sum dp_1 \cdot dp_2 - (\sum dp_1)(\sum dp_2)}{\sqrt{[n \cdot \sum dp_1^2 - (\sum dp_1)^2][n \cdot \sum dp_2^2 - (\sum dp_2)^2]}} \quad (6)$$

From (6), $\gamma(dp_1, dp_2)$ denotes a correlation coefficient, 'n' represents a total number of data packets, dp_1 and dp_2 represents the two data packets, $\sum dp_1 \cdot dp_2$ refers to the sum of cross product of dp_1 and dp_2 , $\sum dp_1$ is the sum of dp_1 score, $\sum dp_2$ is the sum of dp_2 score, $\sum dp_1^2$ is the sum of a squared score of dp_1 and $\sum dp_2^2$ is the sum of a squared score of dp_2 . The correlation coefficient ' γ ' provides the value between +1 and -1, where '+1' denotes a total positive linear correlation, and -1 is a total negative linear correlation between the DP. Based on the correlation measure, the decision dump classifies the DP into different classes. After classifying the DP, the weak learners are combined into one strong learner which is mathematically formulated as follows,

$$H(x) = \sum_{i=1}^n w_i(x) \quad (7)$$

From (7), $H(x)$ denotes an output of the weak learner and $w_i(x)$ denotes a number of weak learners output. Then, the similar weight is assigned to each weak learner.

$$H(x) = \sum_{i=1}^n \alpha * w_i(x) \quad (8)$$

From (8), α denotes a similar weight assigned to weak learners $w_i(x)$. Followed by, the training error of the each base learner is computed as follows,

$$e = (y_i - w_i(x))^2 \quad (9)$$

From (9), e denotes an error. Here, y_i denotes an actual output and $w_i(x)$ denotes an observed output. Based on the error, each base learner is reweighted. Based on the weight value, the boosting classifier finds the classification results with minimum error rate. The weight of the each weak learner is increased if it classified incorrectly. The weight is minimized if it correctly categorized the DP into the different classes. The reweighted classification results are obtained as follows,

$$Y = \sum \alpha^w w_i(x) \quad (10)$$

From (10), Y denotes an output of the strong classification results, α^w reweight of the weak learner $w_i(x)$. Based on the classification results, the DP are correctly classified into different classes. Then the sink node transmits the classified DP into the BS for further processing. The algorithmic process of reweight boosting classification is described as follows,

Input: data packets $dp_1, dp_2, dp_3, \dots, dp_n$, sink node
Output: Improve data classification accuracy
Begin
1. for each data packet dp_i
2. Construct 'n' weak learners $w_n(x_i)$
3. Measure the correlation between $\gamma(dp_1, dp_2)$
4. Classifies dp_i into different classes
5. Combines all weak learners y_i
6. Assign similar weights ' α ' to $w_n(x_i)$
7. Compute the error ' e '
8. Update the weight α^w of $w_n(x_i)$
9. Find classifier $w_n(x_i)$ with $\arg \min e$
10. Obtain strong classification results $Y = \sum \alpha^w w_i(x)$
11. End for
End

Algorithm 3. Reweights boosting classification

Algorithm 3 describes the DP classification using reweight boost ensemble technique. The ensemble technique constructs a number of weak learners. The correlation between the two DP is measured to categorize the DP. Then the weak learners are combined and similar weight is assigned to weak learners. Followed by, the error is computed for each classification results of the weak learner to improve the CA. After that, the weak learner weights are recomputed according to the error value. Then the classifier with minimum error is selected as final strong classification results. By this way, the sink node classifies the DP in an efficient manner and sends to the BS.

4. SIMULATION SETTINGS

The analysis of FGEORBC technique and existing methods [1] and [2] are evaluated in NS2.34 simulator. For data gathering process, totally 500 SN are deployed in the square area of A^2 (1500 m * 1500 m). The Random Waypoint model is used as a mobility model of the SNs. The number of the DP ranges from 10 to 100 for simulation purposes. The simulation time is set as 300 seconds. The Dynamic Source Routing (DSR) protocol is used for energy efficient data gathering in WSN. The table 2 shows the simulation parameters and their values.

Table 2. Simulation Parameters

Simulation parameter	Value
Simulator	NS2 .34
Network area	1500m * 1500m
Number of sensor nodes	50,100,150,200,250,300,350,400,450,500
Data packets	10,20,30,40,50,60,70,80,90,100
Protocol	DSR
Simulation time	300sec
Mobility model	Random Way Point model
Nodes speed	0-20m/s
Number of runs	10

5. RESULTS AND DISCUSSION

The simulation result of FGEORBC technique is compared with existing [1] and [2]. The parameters of FGEORBC are chosen as provided in the experiment section. The results of different FGEORBC technique are discussed using table or graphical representation.

5.1 Simulation Results of Energy Consumption

EC is measured as the amount of energy taken by SN to sense data. EC of SN are computed using the following formula,

$$e_c = \text{no. of sensor nodes} * e(\text{sn}) \quad (11)$$

From (11), e_c denotes a energy consumption where 'e' denotes a energy of the single sensor node (sn).

Table 3. Tabulation for energy consumption

No. of sensor nodes	Energy consumption (joule)		
	FGEORBC	DDGA	MA-Greedy
50	30	40	45
100	35	42	48
150	38	45	54
200	40	46	52
250	43	45	50
300	45	51	56
350	49	56	63
400	50	60	64
450	53	59	63
500	54	61	66

Table 3 describes the simulation results of EC versus number of SN varied from 50 to 500 with three different techniques. The table values clearly show that the energy utilization of the SN is minimized using FGEORBC technique compared to the two existing methods DDGA [1] and MA-Greedy [2]. The minimum EC of FGEORBC technique is achieved by applying fuzzy logic. At first, the entire SN has equal energy which is distributed to the network. Then, these SN senses the environmental conditions. After the sensing behavior of the nodes, the energy of each SN gets minimized. The FGEORBC technique uses the fuzzy logic to verify the SN residual energy with a threshold level. If the energy of the node is greater than the threshold, then the node is considered as greater energy nodes. Otherwise, the node is considered as lesser energy SNs. Lower energy SN has lesser RE and it does not present data aggregation for a longer duration. Thus, FGEORBC technique chooses nodes with greater RE for data aggregation. This helps to minimize EC.

Let us consider the 50 SN for computing the EC. The proposed FGEORBC technique obtains the 30 joules of EC. The EC of DDGA [1] and MA-Greedy [2] are 40joule and 45joule respectively. Similarly, the remaining nine runs are carried out and compare the resultant values of FGEORBC technique and existing methods. Therefore, the FGEORBC technique lessens the EC by 14% and 22% as compared to [1] and [2].

5.2 Simulation Results of Network Lifetime

NL is evaluated as a number of high energy SN is chosen from the total number of SN for data gathering. It is computed as follows,

$$\text{Network lifetime} = \frac{\text{No. of high energy sensor nodes selected}}{\text{No. of sensor nodes}} * 100 \quad (12)$$

The NL of the FGEORBC technique is measured in terms of percentage (%).

Figure 6 shows the results of NL versus number of SNs. The below figure shows that the simulation results of the NL with three different methods which are represented in three different colors namely red, green and blue. Among three different methods, the NL of FGEORBC technique is considerably improved than the two conventional energy-efficient data gathering technique. This improvement of FGEORBC technique is achieved by applying a ranked Gaussian gene optimization algorithm. Initially, 'n' numbers of SN are initialized. Then the low energy SN finds the nearest high energy nodes through the fitness value. If the nodes do not satisfy fitness criteria, then selection and crossover and mutation operations are carried out. After performing the genetic operations, then the fitness is computed to find the higher energy efficient nodes. The selection of high energy nodes improves the NL.

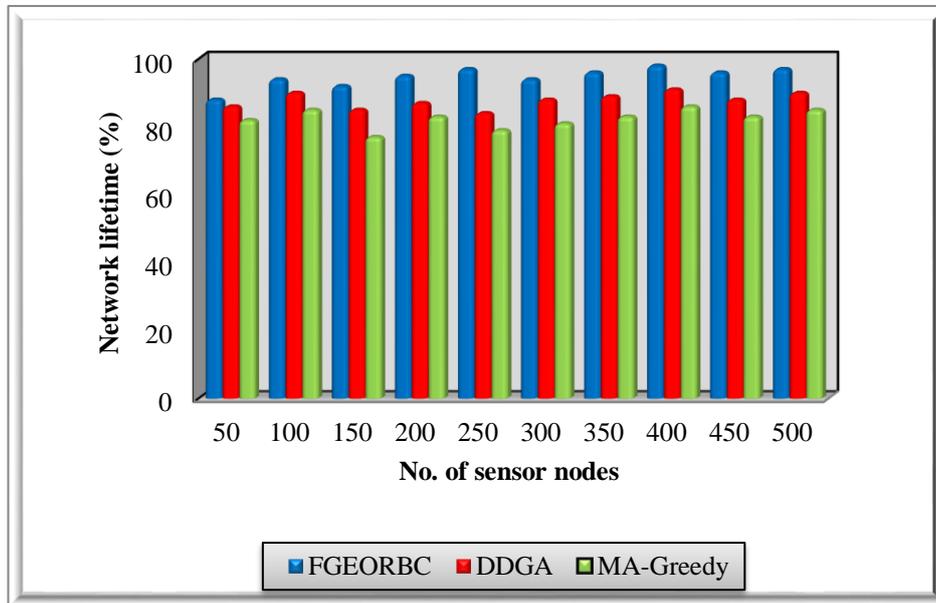


Figure 6. Simulation results of Network lifetime versus the number of Sensor Nodes

Totally ten various results are performed with a different number of SNs. The comparison of the ten different results shows that the FGEORBC technique increases the NL by 9% when compared to existing DDGA [1]. In addition, the NL of FGEORBC technique is improved by 15% compared to existing MA-Greedy [2].

5.3 Simulation Results of Data Gathering Time

Data gathering time is measured as the time taken by the sink node to gather DP higher energy SNs. It is computed as follows,

$$DGT = \text{no. of data packets} * \text{time}(\text{gathering one dp}) \quad (13)$$

From equation (13), *DGT* signify data gathering time, *dp* denotes data packets. The Data gathering time is measured in terms of milliseconds (ms).

Table 4. Tabulation for Data gathering time

No. of data packets	Data gathering time (ms)		
	FGEORBC	DDGA	MA-Greedy
10	17	20	25
20	22	26	30
30	30	33	36
40	32	40	44
50	35	41	45
60	43	46	49
70	46	50	54
80	50	52	55
90	49	53	56
100	56	63	66

Table 4 describes the simulation results of data gathering time with respect to various numbers of DP. The number of DP for the simulation purposes is varied from 10 to 100. The results of data gathering time with three different methods of FGEORBC technique, DDGA [1] and MA-Greedy [2] are described. The results confirm that the proposed FGEORBC technique successfully improves the data gathering with minimum time when compared to existing methods. In WSN, the sink node acts as a data aggregator which collects the data from the high energy SN instead of all the SNs. In an existing conventional approach, the sink node collects the data from all the SN in the network. But, it directs a longer data collection delay. This degrades network performance of networks. In order to resolve this issue, the proposed FGEORBC technique finds higher energy SN and the low energy SNs. The lower energy SN sends their collected data to high energy nodes. The sink node collects the data only from the higher energy SNs. This process improves data gathering efficiency and minimizes the time. While considering 10 DP, the data gathering time of FGEORBC technique is 17ms whereas 20ms and 25ms of time obtained by using DDGA [1] and MA-Greedy [2] respectively. As a result, the data gathering time of FGEORBC technique is significantly minimized by 11% and 19% than the existing DDGA [1] and MA-Greedy [2] respectively.

5.4 Simulation Results of Data Packet Classification Accuracy

Data packet Classification Accuracy is evaluated as the number of DP classified into a total number of DP. It is computed as follows,

$$DPCA = \frac{\text{No. of } dp_s \text{ are correctly classified}}{\text{No. of } dp_s} * 100 \quad (14)$$

From (14), *DPCA* signify data packet classification accuracy, *dp_s* represents the Data Packets. It is measured in terms of percentage (%).

Figure 7 shows the simulation results of data packet CA versus a number of DP.

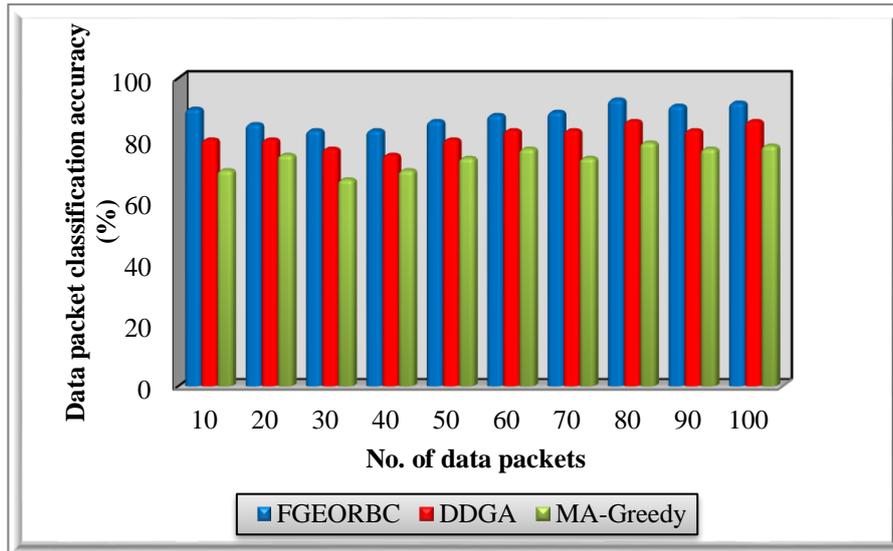


Figure 7. simulation results of data packet classification accuracy versus the number of data packets

The above classification results clearly show that the DP classification results are higher using FGEORBC technique. This is because the sink node collects the DP from the high energy nodes and it categorizes into different classes. The DP classification is performed using an ensemble classifier with the several weak learners. The reweight boosting classifier constructs the decision stump as a weak learner to classify the data based on the correlation between the DP. The boosting technique combines the entire weak learner and similar weight is assigned. The training error for each weak learner is computed. Based on the error value, the weak learners are reweighted. The reweighted results of the boosting classifier effectively categorize the data into different classes with high accuracy. As result, the CA is considerably increased by 8% and 19% when compared to existing DDGA [1] and MA-Greedy [2] respectively.

Based on results and discussion, the FGEORBC technique improves the data gathering efficiency and NL with minimum time as well as high data CA.

6. CONCLUSION

An efficient technique called FGEORBC for improving efficient data gathering with minimum time consumption. The FGEORBC technique includes three processes. At first, the fuzzy logic is applied to determine the higher energy and lower energy SNs. The high RE of the SN minimizes the performance of EC. After that, the low energy SN find the nearest high energy nodes using ranked Gaussian gene optimization. Then the sink node collects the data from the higher energy SNs. This helps to improve the NL. Finally, the sink node performs data classification using reweight boosting technique. The strong classifier efficiently classifies the DP with minimum error. Finally, the sink node sends a classified DP to the base station. The simulation is performed with different metrics EC, NL, data gathering time and CA. The performance result shows that the FGEORBC technique increases the energy efficient data gathering and CA with minimum time as well as EC when compared to the state-of-the-art methods.

REFERENCES

- [1] Yongmin Zhang, Shibo He and Jiming Chen, "Near-Optimal Data Gathering in Rechargeable Sensor Networks with a Mobile Sink", *IEEE Transactions on Mobile Computing*, Volume 16, Issue 6, June 2017, Pages 1718-1729
- [2] Cuicui Lv, Qiang Wang, Wenjie Yan, Yi Shen, "Energy-balanced compressive data gathering in wireless sensor networks", *Journal of Network and Computer Applications*, Elsevier, Volume 61, February 2016, Pages 102-114
- [3] Soobin Lee and Howon Lee, "Energy-Efficient Data Gathering Scheme Based on Broadcast Transmissions in Wireless Sensor Networks", *the ScientificWorld Journal*, Hindawi Publishing Corporation, Volume 2013, July 2013, Pages 1-7
- [4] Prabhudutta Mohanty, Manas Ranjan Kabat, "Energy efficient structure-free data aggregation and delivery in WSN", *Egyptian Informatics Journal*, Elsevier, Volume 17, Issue 3, November 2016, Pages 273-284
- [5] Areej Alsaafin, Ahmed M.Khedr, ZaherAl Aghbari, "Distributed trajectory design for data gathering using the mobile sink in wireless sensor networks", *AEU - International Journal of Electronics and Communications*, Elsevier, Volume 96, 2018, Pages 1-12
- [6] Daisuke Takaishi, Hiroki Nishiyama , Nei Kato , Ryu Miura, "Toward Energy Efficient Big Data Gathering in Densely Distributed Sensor Networks", *IEEE Transactions on Emerging Topics in Computing* , Volume 2, Issue 3, 2014, Pages 388 – 397
- [7] Saud Althunibat , Amir Abu-Al-Aish , Wael F. Abu Shehab , Wael H. Alsawalmeh, "Auction-Based Data Gathering Scheme for Wireless Sensor Networks", *IEEE Communications Letters* , Volume 20, Issue 6, 2016, Pages 1223 – 1226
- [8] Xuan Rao, Hongyu Huang, Jiqiang Tang, Huan Zhao, "Residual energy aware mobile data gathering in wireless sensor networks", *Telecommunication Systems*, Springer, Volume 62, Issue 1, 2016, Pages 31–41
- [9] Mihaela Mitici, Jasper Goseling, Mauritsde Graaf, Richard J.Boucherie, "Energy-efficient data collection in wireless sensor networks with time constraints", *Performance Evaluation*, Elsevier, Volume 102, August 2016, Pages 34-52
- [10] Shouling Ji, Raheem Beyah, and Zhipeng Cai, "Snapshot and Continuous Data Collection in Probabilistic Wireless Sensor Networks", *IEEE Transactions on Mobile Computing*, Volume 13, Issue 3, 2014, Pages 626-637
- [11] Shusen Yang, Usman Adeel, Yad Tahir, and Julie A. McCann, "Practical Opportunistic Data Collection in Wireless Sensor Networks with Mobile Sinks", *IEEE Transactions on Mobile Computing*, Volume 16, Issue 5, May 2017, Pages 1420 – 1433
- [12] Shiliang Xiao, Baoqing Li, Xiaobing Yuan, "Maximizing precision for energy-efficient data aggregation in wireless sensor networks with lossy links", *Ad Hoc Networks*, Elsevier, Volume 26, 2015, Pages 103-113
- [13] Leandro Aparecido Villas, Azzedine Boukerche, Heitor Soares Ramos, Horacio A.B. Fernandes de Oliveira, Regina Borges de Araujo, and Antonio Alfredo Ferreira Loureiro, " DRINA: A Lightweight and Reliable Routing Approach for In-Network Aggregation in Wireless Sensor Networks", *IEEE Transactions on Computers* Volume 62, Issue 4, April 2013, Pages 676–689

- [14] Zhetao Li, Yu Xin Liu, Ming Ma, Anfeng Liu, Xiaozhi Zhang, Gungming Luo, “MSDG: A novel green data gathering scheme for wireless sensor networks”, *Computer Networks*, Elsevier, Volume 142, 2018, Pages 223-239
- [15] Fen Zhou , Zhenzhong Chen , Song Guo , Jie Li, “Maximizing Lifetime of Data-Gathering Trees With Different Aggregation Modes in WSNs”, *IEEE Sensors Journal* , Volume 16, Issue 22, 2016, Pages 8167 – 8177
- [16] Mohamed El Fissaoui, Abderrahim Beni-Hssane, Mostafa Saadi, “Mobile Agent Protocol based energy aware data Aggregation for wireless sensor networks”, *Procedia Computer Science*, Elsevier, Volume 113, 2017, Pages 25-32
- [17] Nidhi Gautam, Sanjeev Sofat, Renu Vig, “Data collection model for energy-efficient wireless sensor networks”, *annals of telecommunications - annales des télé communications*, Springer, Volume 70, Issue 11–12, 2015, Pages 501–511
- [18] Ngoc-Tu Nguyen, Bing-Hong Liu, Van-Trung Phamb, Yi-Sheng Luo, “On maximizing the lifetime for data aggregation in wireless sensor networks using virtual data aggregation trees”, *Computer Networks*, Elsevier, Volume 105, 2016, Pages 99-110
- [19] Divya Lohani and Shirshu Varma, “Energy Efficient Data Aggregation in Mobile Agent-Based Wireless Sensor Network”, *Wireless Personal Communications*, Springer, Volume 89, Issue 4, 2016, Pages 1165–117
- [20] Hedieh Sajedi and Zahra Saadati, “A Hybrid Structure for Data Aggregation in Wireless Sensor Network”, *Journal of Computational Engineering*, Hindawi Publishing Corporation, Volume 2014, December 2014, Pages 1-7

AUTHORS

Mrs. J. Srimathi, Research Scholar in Bharathiar University. Her Field of Interest is Wireless Sensor and Data Base Management System. She has 12 years of Teaching experience, 1 year Software development Experience.



Dr. V. Valli Mayil, Head & Associate Professor, Dept. of Computer Science & Applications, Periyar Maniammai University, Thanjavur. She had Vast Experience in Research Field and she has published more than 20 Research Papers in Articles. Her Interested Research areas are Data Mining and Networks.

