Performance Comparison of Optimization Algorithms for Clustering in Wireless Sensor Networks

N. M. Abdul Latiff  C. C. Tsimenidis  B. S. Sharif
School of Electrical, Electronic and Computer Engineering, Newcastle University, United Kingdom.
abdul-latiff@ncl.ac.uk

Abstract

Clustering in wireless sensor networks (WSNs) is one of the techniques that can expand the lifetime of the whole network through data aggregation at the cluster head. This paper presents performance comparison between Particle Swarm Optimization (PSO) and Genetic Algorithms (GA) with a new cost function that has the objective of simultaneously minimizing the intra-cluster distance and optimizing the energy consumption of the network. Furthermore, a comparison is made with the well known cluster-based protocols developed for WSNs, LEACH (Low-Energy Adaptive Clustering Hierarchy) and LEACH-C, the later being an improved version of LEACH, as well as the traditional K-means clustering algorithm. Simulation results demonstrate that the proposed protocol using PSO algorithm has higher efficiency and can achieve better network lifetime and data delivery at the base station over its comparatives.

1. Introduction

Wireless sensor networks generally consist of a large number of low-cost and low-power sensors that are small in size. Since the energy supply of a sensor node is limited, energy optimization should be considered as the key objective when studying the overall network design problem. Clustering is one of the methods used to manage network energy consumption efficiently. In this approach, each group of sensors has a cluster head node that aggregates data from its respective cluster and sends it towards the base station (BS) as a representative sample of its cluster. The optimal selection of the cluster heads with high energy that scattered evenly in the area can be seen as an NP-hard problem (nondeterministic polynomial time) [1]. Hence, evolutionary algorithms such as PSO and GA are generally more suitable to solve these difficult problems because they are population-based stochastic approaches that can avoid being trapped in a local optimum and can often find a global optimum solution.

Several cluster-based protocols have been proposed in the literature, with the objective of maximizing the sensor network lifetime. The LEACH protocol presented in [2] is a distributed cluster-based protocol in which the cluster heads are selected with some probability. However, LEACH does not guarantee that the desired number of cluster heads is selected and cluster heads are not evenly positioned across the network. In a centralized version of this protocol called LEACH-C [3], a better overall performance of the network is achieved due to improved cluster formation performed by the BS. In [4], the authors used PSO to equalize the number of nodes and candidate cluster heads in each cluster in order to minimize the energy expended by the nodes while maximizing the data transmission. In [5], a protocol that has the objective of minimizing the total sum of the distance squared of all clusters in the network using PSO has been proposed. Nonetheless, the key difference between the proposed work in this paper, the algorithm in [4] and that of [5], is in the application of PSO to choose the optimal nodes as cluster heads to extend the network lifetime.

In this paper, we develop a centralized cluster-based protocol to extend the sensor network lifetime by using PSO and GA algorithms. The main idea in the proposed method is the selection of a cluster head that can minimize the maximum intra-cluster distance between itself and the cluster member, and the optimization of energy management of the network. In next section, we give a brief description of PSO, GA and K-means algorithms. The proposed cluster formation protocol using the optimization algorithms is described in section 3. We present the simulation study of the proposed protocol in section 4, before concluding the paper in section 5.
2. Overview of PSO, GA and K-Means

2.1. Particle Swarm Optimization (PSO)

PSO [6] is an evolutionary computing technique that is based on the principle such as bird flocking. In PSO, a set of potential solutions are called particles that are initialized randomly. During each generation, each particle evaluates its fitness repeatedly until the fitness satisfies the given threshold. The candidate solution in each generation is updated using the following equations:

\[ v_{id}(t) = w v_{id}(t-1) + c_1 \phi_1 (p_{id} - x_{id}(t-1)) + c_2 \phi_2 (p_{gd} - x_{id}(t-1)) \]

\[ x_{id}(t) = x_{id}(t-1) + v_{id}(t) \]  

(1)

Where \( v \) is the particle velocity; \( x \) is the particle position; \( t \) is the number of iterations (generations); \( c_1 \) and \( c_2 \) are two positive constants, called the cognitive and social acceleration factors respectively; \( \phi_1 \) and \( \phi_2 \) are random numbers within the range \([0,1]\) and \( w \) is the inertia weight. Particle’s best position is denoted as \( p_{id} \) while \( p_{gd} \) is the global best.

2.2 Genetic Algorithm (GA)

GA can be viewed as an optimization method based loosely on the Darwinian principles of biological evolution, reproduction and “the survival of the fittest” [7]. In GA, a population consists of a collection of strings called chromosomes, where a chromosome represents different points in the search space. The representation of each chromosome in GA can be in the form of binary strings or real numbers. A fitness or objective function is used to reflect the goodness of each chromosome in the population. GA maintains this population and repeatedly modifies those using methods such as selection, crossover and mutation to produce a new generation of chromosomes. This procedure is repeated until convergence is reached or until a maximum number of generations are achieved.

2.3 K-Means clustering

The K-means algorithm is the most widely used clustering approach that attempts to find the centre point of a cluster by minimizing the distance between points assigned to be within a cluster and at the centre of that cluster [8]. This algorithm begins with a randomly initialize \( K \) cluster centroid points. Then, for each point, K-means assigns the point to the closest cluster centroid before recalculate the cluster centroid vectors of each cluster.

3. Cluster formation

The operation of the proposed protocol is based on a centralized control algorithm that is implemented at the BS, which is a node with a large amount of energy supply. The proposed protocol operates in rounds, where each round begins with a setup phase at which clusters are formed and followed by a steady state. At the starting of each setup phase, all nodes send information about their current energy status and locations to the BS. Based on this information, the BS computes the average energy level of all nodes. To ensure that only nodes with a sufficient energy are selected as cluster heads, the nodes with an energy level above the average are eligible to be a cluster head candidate for this round. Next, the BS runs the optimization algorithm to determine the best \( K \) cluster heads that can minimize the cost function, as defined by:

\[ \cos t = \beta \times f_1 + (1 - \beta) \times f_2 \]  

(2)

\[ f_1 = \max_{k=1,2,...,K} \left\{ \frac{\sum_{\forall n_i \in C_k} d(n_i,CH_k)}{|C_k|} \right\} \]  

(3)

\[ f_2 = \sum_{i=1}^{N} E(n_i) \left/ \sum_{k=1}^{K} E(CH_k) \right. \]  

(4)

Where \( f_1 \) is the maximum average Euclidean distance of nodes to their associated cluster heads, and \( |C_k| \) is the number of nodes that belong to cluster \( C_k \). Function \( f_2 \) is the ratio of total initial energy of all nodes \( n_i \), \( i = 1,2,...,N \) in the network with the total current energy of the cluster head candidates in the current round. The constant \( \beta \) is used to weigh the contribution of each of the sub-objectives. The fitness function defined above has the objective of simultaneously minimizing the intra-cluster distance between nodes and their cluster heads, as quantified by \( f_1 \) and also of optimizing the energy efficiency of the network as quantified by \( f_2 \). According to the cost function defined above, a small value of \( f_1 \) and \( f_2 \) suggests compact clusters with the optimum set of nodes that have sufficient energy to perform the cluster head tasks.

Generally, for a sensor network with \( N \) nodes and \( K \) predetermined number of clusters, the network can be clustered as follows:

1. Initialize \( S \) individuals to contain \( K \) randomly selected cluster heads among the eligible cluster head candidates.
(2) Evaluate the fitness of each individual:
   i. For each node \( n_i, i = 1, 2, \ldots, N \)
      - Calculate distance \( d(n_i, CH_k) \) between node \( n_i \) and all cluster heads \( CH_k \).
      - Assign node \( n_i \) to cluster head \( CH_k \) where:
        \[
        d(n_i, CH_k) = \min_{k=1,2,\ldots,K} \left\{ d(n_i, CH_k) \right\}
        \] (5)
   ii. Calculate the fitness using equation (2) to (4).
(3) Perform the position and velocity update for PSO or perform the genetic operations for GA.
(4) Repeat steps 2 to 4 until the maximum number of iterations is reached.

After the setup phase, the BS transmits the information that contains the cluster head ID for each node back to all nodes in the network. The steady state phase approach is similar to [2] in which the cluster head sets up a Time Division Multiple Access (TDMA) based schedule for its members to avoid collisions among data messages, allowing the radio devices of each member to be turned off at all times, except during their transmission time, to further reduce the energy consumption of the nodes. Once the cluster head finishes receiving data from all of its members at the end of each frame, it performs data fusion and sends the fused data to the BS.

4. Simulations and analysis

The proposed protocol is implemented in Network Simulator (ns2) [9]. We ran the simulations for 100 nodes in a 500 m\( \times \)500 m network area with 20 percent of the total nodes have 5 J of initial energy, while the remaining 80 percent of the nodes have 2 J of initial energy. The BS location was located at (250,575) and we set the number of clusters \( K = 5 \). Throughout the simulations, we considered several random network topologies to get the average results. The radio energy model used in this paper is based on the first order radio model as in [2]. It is assumed that all nodes are stationary and have power control capabilities to vary their transmit power. The simulations continued until all the nodes in the network had consumed all their energy. We set \( \beta = 0.5 \) in (2) to give an equal contribution of each sub-objective.

4.1. Comparison of PSO and GA

In order to ensure a fair comparison between these two algorithms, the size of the swarm for PSO and the population size of GA are set at 30. We employed the Roulette Wheel Selection method [7] in GA that implements the proportional selection strategy. In order to get the best results, we ran several simulations using varied \( c_1, c_2 \) and \( w \) for PSO and varied crossover rates and mutation rates for GA. For PSO, the best values of \( c_1, c_2 \) is 2.0 and \( w \) is 0.72 while for GA, the best crossover rate and mutation rate are 0.70 and 0.01 respectively. Both algorithms are run for maximum generations of 500.

The convergence rate of cost function in (2) with the number of generations for PSO and GA is shown in Figure 1. It is clear that PSO converges at a much faster rate and achieves a better minimum result, compared to that for GA. This proves that PSO is able to give a better global minimum than GA in the context of clustering in WSNs, and yet it is simpler and easier to implement. However, we do not compare the convergence rate of PSO and GA with K-means algorithm in this work because K-means algorithm does not attempt to optimize the energy efficiency in the network, and the objective of K-means algorithm is only to find the most central node in the network.

![Figure 1. Convergence of cost function](image_url)

4.2. Simulation Results

We compared the performance of PSO, GA and K-means algorithms with the well-known cluster-based protocol developed specifically for WSNs, which are LEACH [2] and LEACH-C [3]. Figure 2 illustrates the system’s lifetime, defined by the number of nodes alive over time of the simulation. It can be clearly seen that clustering using PSO, GA and K-Means algorithms can prolong the network lifetime significantly compared to LEACH and LEACH-C. These three algorithms produce better network partitioning with minimum intra-cluster distance, and also cluster heads that are optimally distributed across the network. Thus, the energy consumed by all nodes for communication can be reduced since the distances between member nodes and their cluster heads are shorter. On the contrary, in LEACH and LEACH-C some nodes have to bridge long distances in order to reach a cluster head due to...
poor network clustering. As a result, some nodes dissipate a large amount of energy while transmitting their data to a distant cluster head. However, PSO outperforms GA, since PSO obtains a lower value for the fitness function. Besides, K-means does not consider the nodes’ energies in searching for the cluster centroids, and hence, affects the network lifetime.

The total data messages received at the BS is shown in Figure 3. The plot clearly indicates the effectiveness of the PSO in delivering successfully more data messages than its counterparts. Since the cluster heads need to send the aggregated data to the BS located outside the network area, the selection of cluster heads with higher energy is significant in this case. For this reason, more data messages are sent to the BS using PSO algorithm, because it is unlikely that the cluster head will run out of energy before the steady state phase ends. In contrast, both LEACH and LEACH-C do not guarantee the selection of high energy cluster head. As a consequence, some cluster heads may die at random places and hence, the total data received at the BS is decreased.

5. Conclusions

In this paper we have presented an energy-aware cluster-based protocol for wireless sensor networks using PSO and GA optimization techniques. We have defined a new cost function that takes into account the maximum distance between the non-cluster head node and its cluster head, and the remaining energy of cluster head candidates in the cluster head selection algorithm. It can be observed from the simulations that PSO can converge faster and achieve better global minimum when compared with GA using the same sized population and the same number of maximum generations. Furthermore, PSO algorithm gives a higher network lifetime and delivers more data to the BS compared to GA, K-means, LEACH and LEACH-C. The results also prove that both PSO and GA can give optimal clustering as K-means algorithm, which has been widely used for the clustering problem.

6. References