Volume 5, Issue 2, February 2015 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Clustering in WSN Using Data Mining and Classification Technique

Shruti Kukreja, Prof Poonam Dabas*
Deparment of Computer Science and Engineering, UIET,
Kukrushetra University, Kukrushetra, India

Abstract— A Wireless Sensor Network (WSN) is a network of small sensor nodes which are energy constraint devices and have limited data transmission and computational power. The nodes are arranged in the form of cluster around a Base Station (Sink). Protocol like LEACH defines clustering process in a hierarchical manner in a bid to reduce energy consumption. In this research we propose a refine clustering method to improve the network lifetime. Unlike LEACH where the clustering process is based on random selection of cluster heads (CHs), we use k-nearest neighbour (k-NN) to create cluster in the network then we use local heuristic search technique- tabu search to find CHs. So that minimum energy is consumed in any transmission. We reviewed different techniques for clustering and optimization and finally propose a hybrid method of these two techniques and will later implement the concept in a simulator.

Keywords— Wireless Sensor Network, Sensor Nodes, Base Station, Cluster Head, k-Nearest Neighbour.

I. INTRODUCTION

A Wireless Sensor Network (WSN) consists of small sensor nodes; these sensor nodes are highly energy constraint devices, limited data transmission and have low computational power. These sensor nodes captures information from the environment, processes data and then transmits the data via radio signals [1]. The advantages of WSN are improved coverage, energy efficiency, and superior channel capacity and enhanced target tracking. Generally, WSNs have no fixed structure, and sometimes there is no monitoring station of sensor nodes during the operational life of the network, therefore it is necessary that a WSN must have mechanisms for self-configuration and adaptation in case of failure, insertion or deletion of a sensor node [1].

Sensor nodes are electronic devices, the main components of sensor nodes are units of storage, processing, sensing and transmission. These sensor nodes captures information from the environment, processes data and transmits the data via radio signal. Usually, these nodes have limited battery life due to various constraints such as limited battery life(due to size, cost, etc), low memory, very low data rates, low bandwidth processing, little computing capability and variable link quality. Despite these constraints these constraints, the sensors were deployed in large numbers, they provide with a very real picture of the field being sensed [2]. A sensor network is composed of a huge number of small sensor nodes, which are organized closely either inside the network or very close to it. The sensor nodes, in WSN monitor physical situations such as pressure, sound, temperature etc. and to pass their data over the network to main position.

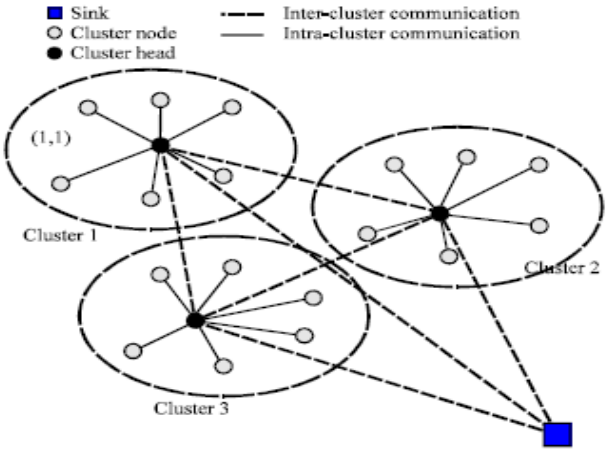

Fig: 1 Data communication in a clustered network.[16]

A. Clustering in WSN

The clustering technique in WSN deals with sensor constraints. The hierarchical network structure allows grouping of sensor nodes into cluster and to assign specific task to the sensor in the cluster, before moving the information to the

higher level. In WSN clustering technique helps to achieve high energy efficiency and assure long network lifetime. In the hierarchical manner (Fig. 1) each cluster has a cluster leader, which is called as the Cluster Head (CH) and several common sensor nodes are the member of the cluster. The cluster formulation process eventually leads to a two level hierarchy where the CH nodes form the higher level and the cluster- member nodes form the lower level [2]. The sensor nodes transmit the data to the corresponding CH nodes. The criteria for selection of the CH are based on the minimum average distance from the base station and residual energy of nodes within the clusters. The CH nodes transmit the data to the base station (BS) either directly or through the intermediate communication with other CH node. In order to balance the energy consumption among all the network nodes, there is rotation (re-clustering) of CH in WSN [3].

B. Clustering Protocol -Low Energy Adaptive Clustering Hierarchy (LEACH)

LEACH s one of the most popular clustering protocol proposed for WSN. It is probability the first dynamic clustering protocol which addresses the need of WSNs. It's a hierarchical, distributed, probabilistic, one-hop protocol, with main objectives (a) to reduce the energy consumption in the network nodes by performing data aggression and thus reducing the number of communication messages (b) to improve the lifetime of the WSN by trying to evenly distribute the energy consumption in the network nodes. [4] It forms cluster based on the received signal strength and uses the CH nodes as router to the BS. All the data processing such as data fusion and aggression are local to the cluster. Distributed algorithm is used by LEACH for cluster formation, where nodes make autonomous decision without any centralized control. Firstly a node decides to be a CH with a probability and then broadcast its decision. Each node non CH node determines its cluster by choosing the CH that can be reached using the least communication energy. The role of CH is rotated periodically among the nodes of the cluster in order to balance the load. And all sensor nodes gets the chance to become CHs.[5] LEACH forms one- hop inter and intra cluster topology where each node can transmit data directly to the CH and then CH transmits the data to the BS. It is mostly applicable to networks deployed in large region.

C. Proposed Clustering Technique –KNN (K nearest neighbor)

KNN classification is one of the most fundamental, simplest and rather trivial classifier in the rote classifier. It memorizes the entire training data and performs classification only if the attribute of the test object match on the training example exactly. *k*NN classification finds a group of *k* objects in the training set that are closest to the test object, and then it bases the assignment of a label on the predominance of a particular class in this neighborhood [6]. There are three key elements of *k*NN approach: the value of *k(* the number of nearest neighbors), a set of labeled objects (e.g., a set of stored records), a distance or similarity metric to compute distance between objects, the number of nearest neighbors. In *k*NN, for classifying an unlabeled object, the distance of the unlabelled object to the labeled objects is computed and its *k*-nearest neighbors are identified. The class labels of these nearest neighbors are used to determine the class label of the object. The drawback of this approach is that many test records will not be classified because they do not exactly match any of the training records.

K-NN classification algorithm:-

The algorithm provides a high-level summary of the nearest-neighbor classification method. Given a training set D and test object $z = (x', y')$ as input. The algorithm computes the distance between z and every training objects, $(x, y) \in D$ in order to determine its nearest neighbor list i.e. D_z. In this algorithm x is the data of a training object, while y is the class. Likewise, $x`$ is the data of the test object and $y`$ is its class [6].

Input: D, the set of k training objects, and test object $z = (x', y')$.

Process:

Compute $d(x', x)$, the distance between z and every object, $(x, y) \in D$.

Select $D_z \subseteq D$, the set of k closest training object to z.

Output: $y' = \text{argmax} \sum_{(xi, yi) \in Dz} I(v = y_i)$.

D. Proposed Optimization Technique – Tabu Search

Tabu search (TS) is a meta-heuristic search method employing local search methods (LS) used for mathematical optimization. TS is a meta-heuristic which means that not guarantee to find an optimal solution, but it is able to find a semi-optimal solution. On the other hand LS finds an optimal solution by checking its immediate neighbours in the hope of finding an improved solution.[7] But LS method have tendency to become stuck in optimal regions where many solutions are equally fit. On the other hand tabu search enhances the performance of LS using these two features. One is compulsory movement to the near solution. It means that when present solution is optimal in the neighbourhood the solution changes to semi-optimal solution. And secondly prohibitions are introduced for changing solution according to the information of a tabu list. TS use memory structure that describes the visited solution and some set of rules. If a potential solution has been previously visited within a certain short- term period or if it has violated a rule, it is marked as "tabu" (forbidden).[7] And the solution search cannot go back because near solutions which have been stored in the tabu list are excluded from the candidate. The properties which exist in tabu list for a certain period discard from this list, because the tabu list has a finite length known as tabu length and it is an important parameter which decides efficiency [11].

The Tabu Search algorithm is as follows :-

Step 1: Choose an initial solution i in S. Set k=0 and i* = i.

Step 2: Set k=k+1 and generate a subset V* of solution in N (I,k) such that either one of the Tabu conditions is violated or at least one of the aspiration conditions holds.

Step 3: Choose a best j in V* and set i=j.
Step 4: If f(i*) > f(i) then set i* = i.
Step 5: Update the Tabu list and aspiration conditions.
Step 6: Check the stopping conditions, if they are met then stop. Else go to Step 2.
Some of the immediate stopping conditions could be the following [4]:

- N (i, K+1) = 0. (no feasible solution in the neighborhood of solution i)
- If the value of K is larger than the maximum number of iterations allowed.
- If the number of iterations from the last improvement of i* is larger than a specified number.
- If an optimum solution has been obtained evidences can be given.

II. RELATED WORK

Plenty of research is available in literature for lifetime prolongation in the field of WSN. Here we are going to review some of the previous studies and researches related to our work of clustering in WSN using different data-mining techniques(*k*-means, EM, PageRank, *k*-nearest neighbour, etc), and optimization techniques (Genetic algorithm (GA), Particle swarm optimization (PSO), Tabu search, etc). Many clustering algorithm in various context have been proposed in the past but to our knowledge, none of these algorithm aims to minimize the energy spent in the system.
In [8], the author discussed the distributed algorithm that organized sensor into hierarchy of cluster with an objective to minimize the total energy spent in the system and communicated the information gathered by the sensor to the information processing centre. But the author failed to explain the underlying medium access protocol and how it would affect the optimal probabilities of becoming a Cluster Head (CH). Jamal and Ahmed in [9] discussed the routing techniques in WSN. These techniques have a common objective i.e. to extend the lifetime of sensor network and not compromising data delivery. But the author failed to integrate the sensor networks with the wired networks. In [7], tabu search and cluster analysis is used. The author had used outage power and outage electricity as the reliability indices, and using these indices the scale of a fault and the time required for the restoration was evaluated. The author had used IEEJ sub-transmission to find an optimal system composition and a restoration procedure. Ameer ahmed and Mohamed in [4], discussed the clustering strategies and algorithm. The author had developed the taxonomy of relevant attributes and had categorized the different schemes according to the objectives, the clustering properties and the clustering process.
In [10], the author had discussed the energy efficient head node selection method for data aggression in WSN by using three aspects i.e. head nodes residual energy that helped head node to prolong its lifetime, the aspect of quality of residual energy helped in balancing the energy consumption for each node in the network and the aspect of total energy spent in the network which helped to save network energy. But the author failed to analyze the energy consumption in different topologies of network. Rhazi, in [11] proposed a scheme to optimize energy consumption in WSN based on Tabu search (TS) optimization technique. The TS adaption consist of defining three types of moves that allowed selecting Cluster head, reassigning nodes to cluster and removing existing cluster. Later the performance was compared to CPLEX- based method, simulated annealing and finally the result showed that TS based resolution method provided a quality solution in terms of execution time and cluster cost.
In [12] Mohammad, used Genetic algorithm (GA) for cluster head selection. The author proposed two algorithms GP-leach and HS-leach to improve the energy consumption by portioning the network and had used evolutionary algorithm for optimized CH selection considering WSN nodes position information and residual energy. GA for Ch selection is good at finding global solution but fails to find local optimum solution. Budha [13] in this research on clustering in wireless network used Particle Swarm Optimization (PSO) for cluster head selection. The study successfully simulated the network and results were more efficient then the base LEACH protocol. But the PSO technique always has early convergence problem. Because only gB best gives the information to others. The PSO evolution only looks for better solution. In [14] performance comparison of four clustering protocol (i.e. SEP, HEED, LEACH and PEGASIS) in WSN for monitoring in terms of network lifetime was done. The author discussed that when simulation id to be done then aspects that should be taken into consideration when designing clustering protocols are the cluster size, selection/rotation of the CH, connectivity and the placement of Base station. But the author failed to study the performance in terms of delay, size of the data sent. J. Hemalatha [15] proposed a scheme based on K-nearest neighbour (KNN) that involved dynamic path selection according to KNN. This scheme improved lifetime by optimized path for the multiple hops between source node and base station.

III. PROBLEM FORMULATED

In a WSN one of the major factor which effect the life of network is the clustering method employed specially in the horizontal routing protocol like LEACH the better clustering and selection of cluster head ensures that node of the nodes need to send data to any node at distance more than a particular optimum value. The maximum energy of nodes is utilized in sending and receiving data packets within network.
Budha [13] in this research on clustering in wireless network used Particle Swarm Optimization (PSO) for cluster head selection. The study successfully simulated the network and results were more efficient then the base LEACH protocol. But the PSO technique always has early convergence problem. Because only gB best gives the information to others. The PSO evolution only looks for better solution. Likewise, Mohammad Karimi [12] used Genetic algorithm (GA) for cluster head selection. GA is good at finding global solutions but fails to find local optimum.
So to overcome this problem we need a optimization technique which gives better global search results. Node of these techniques by Budha and Mohammad ([12], [13]), illustrated the importance of assigning nodes to selective cluster head.

If we use a befitting clustering algorithm for this purpose then the overall cluster creation will be more efficient and consequently network lifetime will improve. In this research, we try to improve the clustering and cluster head selection by applying appropriate classification and optimization technique.

IV. OBJECTIVE

In the last section we explained different researches done in the field of clustering in WSN like the proposed method, most of these work were aimed at improving the lifetime of network by minimizing the energy consumption. We do this by refining the clustering process. The proposed method creates clusters in network and then helps in CH selection to minimize the energy involved in transmitting and receiving data packets.

V. PROPOSED METHODOLOGY

A wireless sensor network consists of large number of small sensor and for clustering these sensors into groups we will use K-nearest neighbor (KNN). KNN will be a helpful technique in creating clusters within the network. This method ensures that no two nodes in the same cluster are at a more than optimum Euclidean distance from cluster centre.

After allocating the nodes to different cluster the next major job is to choice a cluster head in each cluster so that in a single hop protocol like LEACH the cluster head is always located in the best way, so that none of the nodes needs to spent too much energy in sending data to the cluster head.

We first perform the KNN method on the entire set of nodes and then the output of KNN is fed to Tabu Search algorithm which finds the best cluster head from the cluster. Different criteria need to be taken care of while choosing cluster head, like energy level, range and location.

TABLE I NETWORK PARAMETERS TO BE USED FOR SIMULATION

Network node configuration	0.140
MAC	802.11e
Initial energy	0.05J
No of round activity	0:10:120
Simulation time	85sec
Mobility	Random
Routing protocol	/Leach
Speed and data packet size	5m/s,40
Signal	500

Steps for the proposed work :-

Step 1: Take a specific field size to simulate the Wireless sensor network.

Step 2: Locate the sensor nodes in the WSN field in a uniform or a random manner.

Step 3: Accept the value of *k* from the user.

Step 4: Find $n_c = n/k$.

Step 5: Traverse all nodes in network and allocate each node to the nearest key point, so that no key point has more than *k* nodes.

Step 6: Find average distance of each node from its key point, this is the fitness value of this arrangement.

Step 7: update tabu list according to this fitness value to finally find the best location of key points and cluster creation.

Step 8: Use Tabu search Technique to find cluster head for each cluster which is at the least average distance from all nodes in its cluster.

Step 9: Plot the normal and the cluster head nodes in the simulation graph.

Step 10: Perform simulation for certain number of rounds in the WSN where some nodes sends data to Base Station in each round. Calculate energy lost in each round.

Step 11: Compare the total energy lost in available technique to the normal energy loss using basic LEACH protocol.

VI. CONCLUSIONS

This research presented a hybrid methodology to perform clustering. LEACH- the clustering protocol randomly selects CH and then assigns nodes to CHs. We refine this process by using KNN clustering technique to create clusters and then used a local search optimization technique i.e. tabu search for selection of cluster head. The selection is optimized to minimize energy consumption. The proposed method very effectively increases the lifetime of network in comparison to the base of this research [13] by Budha.

The current search is aimed at improving the clustering methodology to maximize the network lifetime. But it is difficult to consider the energy conservation aspect and simultaneously taking care of network security. The future researcher can work on these two aspects together and propose a amendment or may be an entirely new protocol which satisfies both the front.

ACKNOWLEDGMENT

With the overwhelming sense of respect, I would like to convey my heart-felt gratitude to Mrs. Poonam Dabas (Asstt. Proff.), for continuous encouragement and support she provided me during the tenure of my work at UIET, Kurukshetra university, Kurukshetra, and giving me the opportunity to carry out my work in this organization.

REFERENCES

[1] Gustavo S. Quirino, Admilson R. L. Ribeiro and Edward David Moreno, *Asymmetric Encryption in Wireless Sensor Networks*, ISBN 978-953-51-0735-4, September 2012.

[2] L.M.C. Arboleda and N. Nasser, *Comparison of Clustering Algorithms and Protocols for Wireless Sensor Networks*, Proc. Canadian Conf. Electrical and Computer Eng, IEEE, pp. 1787-1792, May 2006.

[3] Raju Duttat, Shishir Gupta and Mukul Kumar Das, *Mathematical Modelling of Packet Transmission through Cluster Head from Unequal Clusters in WSN*, 2012 2nd IEEE International Conference on Parallel, Distributed and Grid Computing, pp.515-520, Dec 2012.

[4] Ameer Ahmed Abbasi and Mohamed Younis, *A survey on clustering algorithms for wireless sensor networks*, in Science Direct Computer Communications, pp. 2826–2841, June 2007.

[5] Basilis Mamalis, Damianos Gavalas, *Clustering in Wireless Sensor Networks*, chapter 12, Zhang/RFID and Sensor Networks, pp. 324-333, 2009.

[6] XindongWu , Vipin Kumar ,J. Ross Quinlan and Joydeep Ghosh, *Top 10 algorithms in data mining*, Springer-Verlag London, Vol. 14, Issue 1, pp 1-37, Dec 2007.

[7] Eisuke Shimoda, Hiroshi Ohtake, Shinichi Iwamoto and Masayuki Namba *Reliability Evaluation of Multiple Facility Faults for Sub-Transmission Systems Using Tabu Search and Cluster Analysis*, Power Engineering Society General Meeting IEEE, 2006.

[8] Seema Bandyopadhyay and Edward J. Coyle *An Energy Efficient Hierarchical Clustering Algorithm for Wireless Sensor Networks*, 22nd Joint Conference of the IEEE Computer and Communications. IEEE, Vol.3, pp. 1713-1723, 2003.

[9] Jamal N. Al-Karaki and Ahmed E. Kamal, *Routing techniques in wireless sensor networks: A survey*, IEEE Wireless Communications", Vol. 11, Issue. 6, pp. 6-28, December 2004.

[10] Wanming Chen, Max Q.-H. Meng, Shuai Li, Tao Mei, Huawei Liang and Yangming Li, *Energy Efficient Head Node Selection Algorithm in Wireless Sensor Networks*, IEEE International Conference on Robotics and Biomimetics, pp.1366-1371, Dec. 2007.

[11] Wanming Chen, Max Q.-H. Meng, Shuai Li, Tao Mei, Huawei Liang and Yangming Li, *Energy Efficient Head Node Selection Algorithm in Wireless Sensor Networks*, IEEE International Conference on Robotics and Biomimetics, pp.1366-1371, Dec. 2007.

[12] Mohammad Karimi, Hamid Reza Naji, and Shahrzad Golestani, *Optimizing Cluster-Head selection in Wireless Sensor Networks using Genetic Algorithm and Harmony Search Algorithm*, 20th Iranian conference on Electrical Engineering (ICEE) , pp. 706-710, 2012.

[13] Buddha Singh and Daya Krishan Lobiyal, *A novel energy-aware cluster head selection based on particle swarm optimization for wireless sensor networks*, Springer, July 2012.

[14] Mariam Alnuaimi, Khaled Shuaib, Klaithem Al Nuaimi and Mohammed Abdel-Hafez, *Performance Analysis of Clustering Protocols in WSN*, WMNC 2013 6th joint IFIP, IEEE, pp 1-6, April 2013.

[15] J. Hemalatha, R. PalaniKumar and Dr. K. Ramasamy, *KNN Based Power Efficiency in Wireless Sensor Networks*, in International Journal of Emerging Technology and Advanced Engineering**,** ISSN 2250-2459, ISO 9001:2008 Certified Journal, Vol. 3, Issue 4, pp. 347-350, April 2013.

[16] Chao Gao, Xiaoya Hu, Bingwen Wang, Hongliang Gao and Wei Xiong, *Data Dissemination in Wireless Sensor Networks with Clustering Method*, Information Technology Journal, vol. 11, pp. 1477-1483,2012.

