

Data Relay Clustering Algorithm for Wireless Sensor Networks: A Data Mining Approach

¹S. Nithyakalyani and ²S. Suresh Kumar

¹Department of Computer Science and Engineering,
K.S.R College of Engineering, Tamilnadu, India

²Vivekanandha College of Technology for Women, Tamilnadu, India

Abstract: Problem statement: Nowadays sensors are very essential for today life to monitor environment where human cannot get involved very often. Wireless Sensor Networks (WSN) are used in many real world applications like environmental monitoring, traffic control, trajectory monitoring. It is more challenging for sensor network to sense and collect a large amount of data which are continuous over time, which in turn need to be forwarded to sink for further decision making process. Clustering of sensory data act as a nucleus job of data mining. A clustering in WSN involves selecting cluster heads and assigning cluster members(sensors) to it for efficient data relay. The constraints in power supply, limited communication, bandwidth, storage resources are the major challenges in WSN facing today. **Conclusion:** Proposed study presents K-Means Data Relay (K-MDR) clustering algorithm for grouping sensor nodes there by reducing number of nodes transmitting data to sink node, it reduces the communication overhead and in this manner increase the network performance. Furthermore Conserve and Observe Modes (COM) algorithm reduces the number of nodes within the cluster there by without compromising the coverage face major challenges such as limited communication bandwidth, constraints in power supply and storage resources region of it. The contribution of K-MDR is to reduce power consumption finally the simulation experimental results show that the time efficiency of the algorithm is achieved.

Key words: WSN, clustering, COM, data mining

INTRODUCTION

Advances in wireless communications made to cultivate tiny hardware components as multifunctional and intelligent sensor nodes with major advantage of low-Power and low-cost. Usually it communicate in short range distances over a radio frequency channel and these devices are small in size. The components of these tiny nodes are sensing, processing and communicating data, realize the objectives of wireless sensor networks (Taherkordi *et al.*, 2008).

Comparing with the tradition sensors, Wireless sensor networks promise significant improvements. A large number of integrated sensor nodes from the Wireless Sensor Network which are densely deployed either inside the observable fact or very close to it. It cooperate with each other through a wireless network in gathering the environmental information or reacting to particular events. Classical applications of sensor networks are monitoring of medical data, weather monitoring, object tracking, vehicle monitoring and combat field survey (Ilyas and Mahgoub, 2005). The

majority of sensor networks applications fall into the querying class of applications and for future analysis and mining it required to continuously collect and integrate data. The WSN's extraordinary characteristics direct us to innovative research challenges in several data mining process. WSN face rigorous resource constraints in communication bandwidth, power supply, storage and processor capacity (Ma *et al.*, 2005) normally the traditional mining techniques is centralized, computationally expensive and focuses on disk stored data. In data mining grouping a similar data is known as clustering which is a preparatory step for future data analysis. In this study a new algorithm called distributed K-means clustering algorithm is proposed for clustering sensor's node in WSN. The nodes within the clusters will forward the data to sink through cluster head where aggregation of data takes place.

Cluster analysis in data mining: The process of clustering the data objects into cluster is depended on the similarities and their functionalities is data clustering. This process is much easier to collect and

Corresponding Author: S. Nithyakalyani, Department of Computer Science and Engineering, KSR College of Engineering, Tiruchengode, India

process the data sensed from the environment. The dissimilar sensors are placed in the nearby clusters. The difference between the clusters are dependent on the characteristics of each cluster and they are measured by various distance functions like Manhattan distance and Eulidean distance (Han *et al.*, 2011; Forman and Zhang, 2000). There are number of algorithms are developed for data clustering in the past. Many of these clustering algorithms are designed to deal with data which specifically stored in a traditional database management system, but here WSN is distributed environment so the data taken to analyze is ditributed across multiple sites. The distributed environment faces many challenges in data analysis because of privacy and limited bandwidth (Silva *et al.*, 2005). In distributed environments many algorithms are proposed for the problems of cluster analysis. Silva *et al.* (2005), based on rounds of message passing between data sites and a central site a number of distributed algorithms are grouped. A wireless sensor network is made up of energy efficient sensors in a distributed environment. Therefore, these challenges should be addressed while designing a data clustering algorithms and also the sensor constraints in communication and computation are considered.

Cluster support sensor networks: The energy limitations in sensor network makes hard in managing the sensor energy and safeguarding of the power is a crucial factor for attaining prolonged network lifetime. Clustering sensors into group is more efficient and adaptive approach is adopted for node communication and routing by many studies on sensors energy consideration (Kim *et al.*, 2005).

Cluster-based mechanism (Ghiasi *et al.*, 2002) is employed for node communication and routing. In the clustered WSN the sensors communicate data only to cluster-head. Then the cluster heads transfer the aggregated data to the processing center or to the base station which is called as sink shown in Fig. 1. The base station is a specialized device or one of the sensors.

Bandyopadhyay and Coyle (2003) selecting a set of cluster head among the nodes in the network is the essential operation, it also cluster the other nodes with these heads, the cluster heads are selected according to some negotiated rules, normally the node which is more powerful in the topology play the role of cluster and the other nodes will forwarded the sensing data to cluster nodes. In energy consumption sensors, communication is the main factor to be considered.

Ghiasi *et al.* (2002) the amount of energy used is totally dependent on the distance the data has to travel between the sender and the receiver. Since the sensors only communicate data to cluster heads over smaller distances in clustered sensor network, the total number of sensors in the network is much lower than the situation in which each and every sensor communicates directly to the base station.

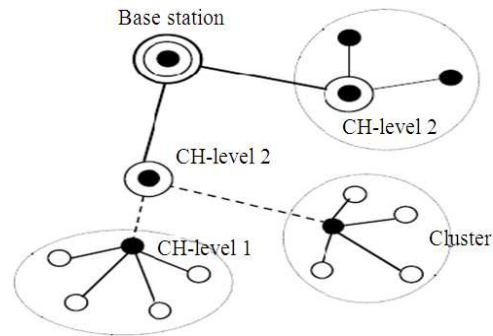


Fig. 1: Cluster -based Sensor network

Lee *et al.* (2004) many heuristic algorithms have been proposed for choosing cluster-heads. Moreover, in all the approaches, the attempt is made to estimate the optimal number of cluster-heads.

Kim *et al.* (2005) to balance the energy consumption over the entire network is the objective of WSN. The algorithms like ASCENT, GAF, LEACH, SPAM, HEED and ACE are all try to preserve and balance the energy dissipation of the WSN using cluster based architectures.

The major goal of clustered based sensor network is to stability the energy consumption over the entire network. LEACH, ASCENT, SPAN, GAF, ACE and HEED all attempts to save and balance the energy dissipation of the network using cluster-based architectures, as a result extending entire the network lifetime (Lee *et al.*, 2004).

Problem definition: Consider N number of sensor node is dispersed uniformly in a field to detect D attributes. Here the assumption is with predetermined cluster-heads these sensors are clustered into groups resulting in a clustered sensor network. In our proposed approach the goal is to develop an algorithm that clusters the the sensor nodes for a data relay without compromising coverage area. The majority of the algorithms are not concentrated on the convergence time and computation time.

Proposed work: Initial assumptions made for following work:

- The network is homogeneous
- For long range transmissions the cluster-heads are dominant for performing computations to the sink
- Each and every sensor node transfers data directly only with other nodes in its cluster
- Number iteration in K-means controlled by sum of mean square error

Generally K-means algorithm is designed to group the sensors nodes for proficient data relay. First and foremost, k random points are chosen as the cluster leader by the sink node. The main drawbacks of the above mentioned algorithm are that the chosen cluster heads may not be able to obey its assigned function as a representative. Moreover in case of smaller networks it would result in a high cost for clustering the predefined number of k, whereas the clusters could be limited. The proposed K-MDR algorithm conserves the energy by delimiting the clustering factors in a number of ways and thus improving efficiency.

K-MDR Algorithm

- 1: Let select k random points as $Ch_j(j= 1,2,3,4,..,k)$
- 2: repeat
- 3: for all $s_i \in S$ do /*assign each sensor to clusters*/
- 4: $Q_i \leftarrow Ch$ /*candidate clusters*/
- 5: to prune candidate representatives from
 $Ch = \{ch_1, ch_2, ch_3, \dots, ch_k\}$.
- 6: for all $ch_j \in Ch$ do
- 7: if $MDR_i \subseteq V(ch_j)$ then
- 8: $Q_k \leftarrow \{ch_j\}$ /*The one and only one candidate*/
- 9: if $|Q_i| = 1$ then /*only one candidate remains*/
- 10: $f(i) \leftarrow j$ where $ch_j \in Q_i$
- 11: else
- 12: for all $ch_j \in Q_i$ do /*remaining candidates*/
- 13: Compute $ED(s_i, ch_j)$
- 14: $f(i) \leftarrow \text{argmin}_j: ch_j \in Q_i \{ED(s_i, ch_j)\}$
- 15: for all $j=1, \dots, k$ do /*readjust cluster representatives*/
- 16: $ch_j \leftarrow$ centroid of $\{s_i \in S \mid f(i) = j\}$
- 17: until Ch and f become stable $f(i) \leftarrow j$ *

For improving the performance of K-means algorithm, the time spent on ED calculations need to be reduced, because it dominates the execution time of the algorithm. Since we know that computation cost is much more than transmission cost. The one of the possible way to achieve is whenever possible ED computation can be avoided. To the specified sensor s_i , the set Q_i supplies the group of candidate cluster representatives that are likely the closest to s_i . Initially, $Q_i = Ch$, a pruning algorithm is useful to prune candidate representatives from Q_i that are definite to be not nearby to sensor s_i . Suppose if one candidate cluster leftover in Q_i , then sensor s_i is hand overed to that cluster. If not, calculate the expected distances between s_i and other remaining clusters in Q_i . Sensors with the smallest expected distance are assigned to the cluster. A excellent pruning algorithm should supposed to considerably reduce the set Q_i to a very small cardinality, so that number

of ED calculations that required to be carry out in line 13 is as not many as possible. Thus we can reduce ED calculation by K-MDR.

Above method would swallow a considerable amount of energy of the individual sensor in the network. Additionally, another algorithm (COM) would render its function to make the lifetime of the entire network to be enhanced without comprising coverage area within its cluster by minimizing number of nodes contributing to sense and forward data.

COM Algorithm for sensor activation: After the first iterative step, the further selected subset of nodes enters into the monitoring state while the remaining sensor nodes go to conserve mode. The selected observe nodes provide full coverage over the monitored field during this iteration. In COM, an activation delay is assigned by each sensor node themselves that is relative to its function cost $C(s) = 1/E(s)$. For choosing the active sensors the next iteration in this algorithm a smallest cost sensor node have a highest chance of becoming active sensors. Every sensor node then waits for a amount of time before come to a decision whether to stay in observe mode in the next communication round. Until sensor wait for a observe mode delay time to expire, the sensor node can accept the observe mode communication from its neighboring nodes, which have smaller activation delays (smaller cost), if they decide to become active during the upcoming iteration. Once the sensor node's observe mode delay time expires, the sensor node check whether the neighboring nodes covers its region entirely for monitoring process. If not, the sensor node broadcast an observe mode message to its neighboring nodes about its decision to stay active. Nodes with the minimum cost have more priority to become active. In the network all sensor nodes mutually take part in the activation phase, despite of the cluster to which they fit in. This eliminates the redundant activation of sensor nodes on the borders of the clusters, which may happen when the activation of nodes is done in each cluster independently.

Experimental Result: The simulation results of the proposed work are carryout with MATLAB. And comparison between K-MDR and K-MEANS algorithm (Fig. 2) are given. From the results obtained, we find that K-MDR with COM was proved to be effective and efficient algorithm for node clustering (Fig. 3). Furthermore optimal number of nodes was participated per iteration in clusters without compromising network coverage.

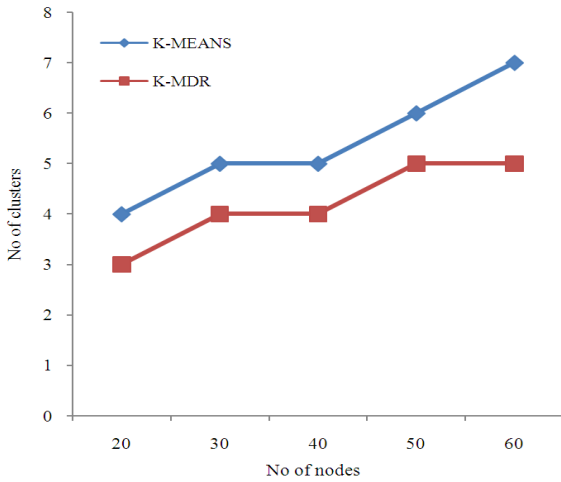


Fig. 2: No of clusters

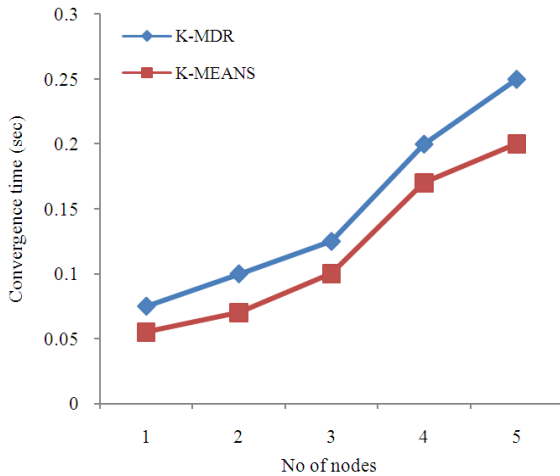


Fig. 3: Convergence diagram

CONCLUSION

Thus the proposed K-Means Data Relay (K-MDR) clustering algorithm for WSN reduces the communication overhead and increases the entire network life time by reducing the number of transmission between every sensor node to sink. The K-Means Data Relay algorithm decreases the computational time and improves the performance of the network when compared to K-Means algorithm. Furthermore, Conserve and Observe Modes (COM) algorithm reduces the number of nodes within the cluster thereby without compromising the coverage region for controlling the energy of network. Energy efficiency of nodes is improved by limiting their energy consumption.

REFERENCES

- Bandyopadhyay, S. and E.J. Coyle, 2003. An energy efficient hierarchical clustering algorithm for wireless sensor networks. Purdue University.
- Forman, G. and B. Zhang, 2000. Distributed data clustering can be efficient and exact. ACM SIGKDD Exp. Newsletter, 2: 34-38. DOI: 10.1145/380995.381010
- Ghiasi, S., A. Srivastava, X. Yang and M. Sarrafzadeh, 2002. Optimal energy aware clustering in sensor networks. Sensors, 2: 258-269.
- Han, J., M. Kamber and J. Pei, 2011. Data Mining: Concepts and Techniques. 3rd Edn., Elsevier, Amsterdam, ISBN-10: 9780123814791, pp: 744.
- Ilyas, M. and I. Mahgoub, 2005. Handbook of Sensor Networks: Compact Wireless and Wired Sensing Systems. 1st Edn., CRC Press, Boca Raton, ISBN-10: 0849319684, pp: 864.
- Kim, H., S.W. Kim, S. Lee and B. Son, 2005. Estimation of the optimal number of cluster-heads in sensor network. Knowl. Based Intell. Inform. Eng. Syst., 3683: 181-181. DOI: 10.1007/11553939_13
- Lee, S.H., J.J. Yoo and T.C. Chung, 2004. Distance-based energy efficient clustering for wireless sensor networks. Proceedings of the 29th Annual IEEE International Conference on Local Computer Networks, Nov. 16-18, IEEE Xplore Press, pp: 567-568. DOI: 10.1109/LCN.2004.44
- Ma, X., D. Yang, S. Tang, Q. Luo and D. Zhang *et al.*, 2005. Online mining in sensor networks. Lecture Notes Comput. Sci., 3222: 544-550. DOI: 10.1007/978-3-540-30141-7_81
- Silva, J.C.D., C. Giannella, P. Bhargava, H. Kargupta and M. Klusch, 2005. Distributed data mining and agents. Eng. Appli. Artif. Intell., 8: 791-807. DOI: 10.1016/j.engappai.2005.06.004
- Taherkordi, A., R. Mohammadi and F. Eliassen, 2008. A communication-efficient distributed clustering algorithm for sensor networks. Proceedings of the 22nd International Conference on Advanced Information Networking and Applications, Mar. 25-28, IEEE Xplore Press, Okinawa, pp: 634-638. DOI: 10.1109/WAINA.2008.130