International Journal of Engineering Science Invention (IJESI)
ISSN (Online): 2319 – 6734, ISSN (Print): 2319 – 6726
www.ijesi.org || PP. 230-235

Analytical Approach Based on Comparison between K-meansand KNN Algorithm for Clustering and Data Aggregation in WSN

Seema Taranum[1], Nalina S.B[2]
(Department of ECE, JNNCE, India)(Department of ECE, JNNCE, India)

Abstract: *Wireless sensor network (WSN) is one of the Ultimate technology and are resource constrained devices. Clustering and Data Aggregations are two important factors of WSN which is used to reduce the Energy Consumption in the network by decreasing the amount of Data Transmission. Machine Learning algorithms such as Supervised, Unsupervised and reinforcement learning, neural networks significantly reduce the amount of data transmission and use the distribution characteristics of the network. It provides a comparative analysis of the performance of different methods to help the designers for designing appropriate machine learning based solutions for clustering and data aggregation applications. This paper presents a literature review of different machine learning based methods which are used for clustering and data aggregation in WSN and proposes an improved similarity based clustering and data aggregation, which uses Independent Component Analysis (ICA).*

Keywords: *WSN,Clustering, Data aggregation,ICA, Energy Consumption, Network lifetime, Delay*

I Introduction

Wireless sensor networks (WSNs) concern networks of spatially scatter and dedicated sensors that monitor and record the material environment and forward the composed information in visible form. WSNs can measure tangible environment such as hotness, sound, contamination levels, very damp weather and wind. WSNs monitor tangible or tangible environment, such as coldness of some degree, sound, and pressure. Modern networks happen bi-directional; two together accumulate information in visible form and enabling control of sensor special interest or pursuit. The incident of these networks was stimulate by military hard work to a degree battlefield following.

The earth's features of a WSN can vary from a plain star network to a manner multi-hop communicating without material contact mesh network. Propagation can give money in exchange for work performed send along a path or flooding[1]. Machine Learning (ML) does the process that certainly improves the study or occurrence, and acts beyond being specifically prioritize. ML happen making our computing processes effective, trustworthy and cost-effective. ML produce models by break down to components even complicated information in visible form automatically, fast and exactly. It happen mainly confidential into directed learning, alone knowledge, to a certain extent-supervised knowledge and military aid learning. The substance of ML display or take public their power to act to provide make a sweeping assumption resolution through an structure of something that can learn to make or become better allure performance. Because of the combining two or more academic field's character, it plays an important role fashionable miscellaneous fields including the act of turning material to use, healing, and calculate. Recent advances in ML bear exist applied to answer differing issues fashionable WSNs. ApplyingML not only improves the acting of WSNs and in addition to limits the human the act of intervening or re- program. Access wide in range amount of information in visible form collected for one sensors, and extract the beneficial news from the data happen not so smooth without ML. It in addition to uses to merge Internet of belongings (IoT), cyber-concerning the body arrangement (CPS) and machine to device that performs a task (M2M). Energy-gathering give a self-powered and long-lasting protection for the WSNs redistribute with inside the bitter atmosphere. Sensor knot may in addition to alternate their extent by way of a few central or outside determinant. Accurate localization exist easy and fast accompanying the assist of ML algorithms. ML is used to discriminate and separate the not working sensor nodes from common sensor growth and improve the performance of the society. Transmitting all news to the bottom station will influence broadcast overhead accompanying inside the community. ML furthermore admit lessening the range of the information ahead of the sensor or group head.

II Clustering and Data AggregationClustering

Clustering in Wi-Fi sensor networks, Sensor nodes in surroundings gather facts and transmit it to a sink both at once or collaboratively via different nodes. Many sensor packages cluster the sensor nodes to acquire scalability,robustness and decreased community traffic. Clustering answer can advocate a sleep/wakeup agenda for a WSN to efficiently lessen strength intake. In many sensor applications, all of the sensor hubs aren't required to be in wakeup mode and consume power. Based at the worldly and spatial need, a few sensor growth grant permission be set up at sleep mode in what way no power stay consumed. In clustering, the sensor nodes are partitioned intospecific clusters. Each cluster is controlled via way of means of a node referred as cluster head (CH) and different nodes are referred as cluster nodes. Cluster nodes do now no longer speak at once with the sink node. They must by skip the amassed information to the cluster head. Cluster head gathers the information, obtained from cluster nodes and transmits it to the bottom station. Thus minimizes the power intake and variety of messages communicated to base station. Also variety of energetic nodes in conversation is reduced. Ultimate end result of clustering the sensor hub is extended community lifetime. [7]

Sensor Node: It is the core component of wireless sensor network. It bears the capability of subject to series of actions to achieve result, etc.

Cluster Head: The Cluster head (CH) is taken into consideration as a pacesetter for that unique cluster. And it's far chargeable for exclusive sports completed with inside the cluster, inclusive of records aggregation, records transmission to base station, scheduling in the cluster, etc.

Base Station: Base station is taken into consideration as a first-rate records series node for the whole sensor community. It is the bridge (thru verbal exchange link) among the sensor community and the stop user. Normally this node is taken into consideration as a node without a energy constraints.

Cluster: It is the organizational unit of the community, created to simplify the verbal exchange with inside the sensor community.

DATA AGGREGATION IN A CLUSTERED ARCHITECTURE

Another fundamental issue that contributes to energy-performance in a clustered structure is using facts aggregation for energy-green facts collection. Urban areas like homes and rooms endowment a plot, in which it form better feel to consolidation inside information that belongs to a cluster of sensors placed fashionable the boundaries of a authorize field. It's bothersome to please expansive sums of news honestly to the understand or be understood a WSN. The aggregator hub takes information from a few sensor nodes, aggregates it utilizing conglomeration capacities (such as Number, MAX, Whole, and MIN), and after that sends it to the sink hub. By bringing down the number of bundle transmissions, this strategy evacuates duplication within the obtained information and increments arrange life time. As a result of the aggregation process, the number of packets and collisions, as well as the number of retransmissions, are reduced. A lower number of retransmissions reduce power and time waste while also increasing network throughput. [10]

Functional Challenges Clustering and Data Aggregation

Clustering & data accumulation basically comes in WSN, ML methods progress the operation of hub clustering and information conglomeration as takes after

- Utilization of machine learning to compress information locally at cluster heads by efficiently extracting similitude and divergence (e.g., from flawed hubs) in different sensors readings.
- Machine learning calculations are utilized to efficiently choose the cluster head, where suitable cluster head choice will significantly diminish vitality utilization and improve the network’s lifetime

K-NEAREST NEIGHBOR (K-NN)

This supervised gaining knowledge of set of rules classifies a records sample (known as a question point) primarily based totally at the labels (i.e., the output values) of the close to records samples. For example, lacking readings of a sensor node may be anticipated the usage of the common measurements of neighboring sensors inside precise diameter limits. There are numerous features to decide the closest set of nodes. A easy approach is to apply the Euclidean distance among one-of-a-kind sensors. K-nearest neighbor does now no longer want excessive computational power, because the characteristic is computed relative to nearby points (i.e., ok-nearest points, in which ok is a small nice integer). This issue coupled with the correlated readings of neighboring nodes makes ok-nearest neighbor an appropriate disbursed gaining knowledge of set of rules for WSNs. In [38], it's been proven that the ok-NN set of rules may also offer erroneous effects whilst studying issues with excessive- dimensional spaces (extra than 10-15 dimensions) as the space to one-of-a-kind records samples turns into invariant (i.e., the distances to the closest and farthest friends are barely similar). In WSNs, the maximum critical utility of the ok-nearest neighbor set of rules is with inside the question processing

subsystem

K- MEANS:

The k-approach set of rules is used to apprehend statistics into distinct classes (referred to as clusters). This unsupervised gaining knowledge of set of rules is extensively utilized in sensor node clustering trouble because of its linear complexity and easy implementation. The k-approach steps to solve such node clustering trouble are

(a) Randomly choose k nodes to be the preliminary centroids for distinct clusters;

(b) Label every node with the nearest centroid the usage of a distance function;
(c) Re-compute the centroids the usage of the modern node memberships and
(d) Prevent if the convergence situation is valid (e.g., a predefined threshold for the sum of distances amongnodes and their attitude centroids), in any other case pass lower back to step (b).

INDEPENDENT COMPONENT ANALYSIS

Independent Component Analysis (ICA) is a machine learning technique to separate independent sources from amixed signal unlike most important thing evaluation which makes a Speciality of maximizing the variance of the information points; the unbiased thing evaluation makes a speciality of independence, i.e. unbiased components.

IV. Proposed Method

Clustering and Data Transmission Using Machine Learning

Cluster heads are selected using two factors in conventional sensor network clustering methods: based on residual and duration from the base station's node, however cluster heads are selected using a cost metrics function in machine learning-based clustering. In the usual approach of WSN clustering, there are overloaded cluster heads, however in sparse areas, there may be very few cluster heads or no cluster heads at all.

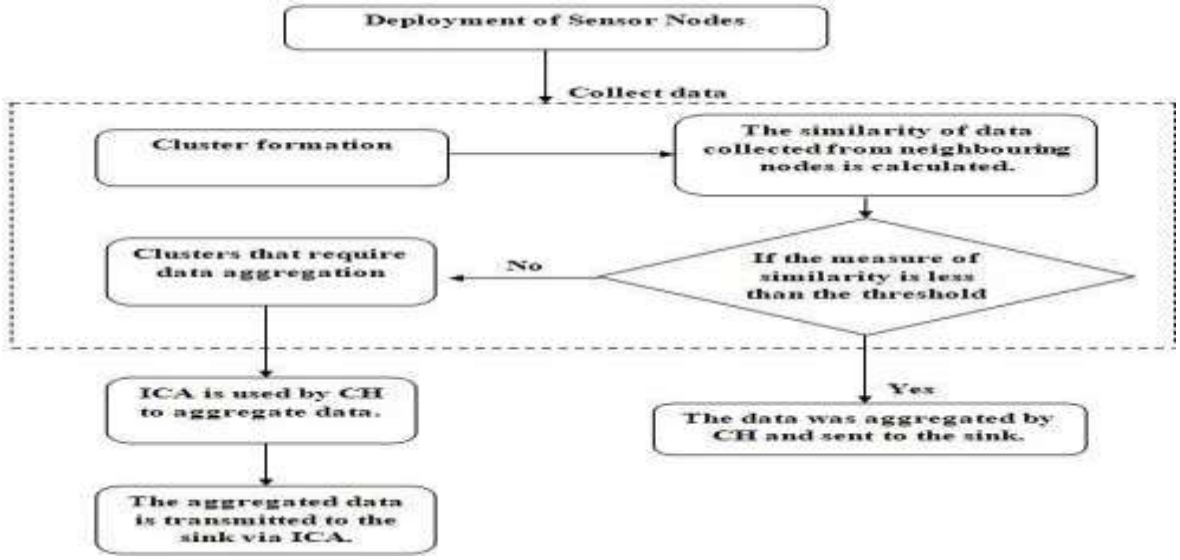


Fig 1: Proposed Flow Diagram for the data aggregation phase

This paper proposes the Independent Component Analysis (ICA) methodology for cluster-based data aggregation, which is used on similar data clusters. The CH node's ICA. In comparison to other techniques such as PCA and compression algorithms, the ICA method performs well. It employs the concept of differential entropy to reduce mutual information.. The sink node receives consolidated data from similar clusters. As a result of the reduced number of aggregating procedures, the computation and energy consumption maybe reduce. The CH does the information in visible form collection. For information in visible form reduction, ICA happen computationally persuasive.

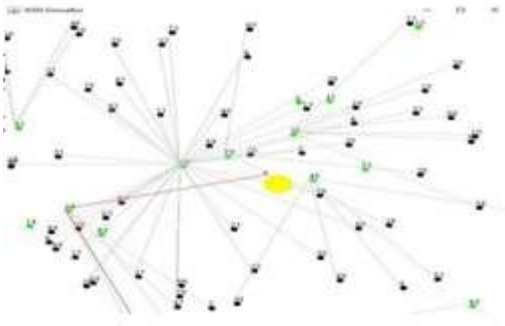


Figure 2: Data transmission and aggregation

Figure 3: Data transmission and aggregation between nodes using K-means algorithm.Bluenodes serve as supplementary nodes, which helps in lowering energy consumption.

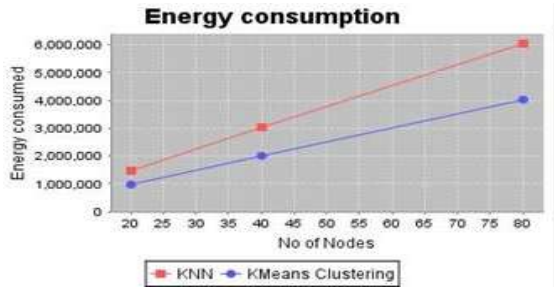
V. Simulation Results

Net bean was employed in this dissertation. Net Bean is a java-based IDE that assists in the creation and execution of java-based projects. All nodes in the simulation communicates with data from their present position to the cluster head node, which then delivers it to the base station. This information will assist us in creating a graph for these factors. The output diagrams below show are the results achieved. The performance analysis for these parameters can be obtained from the output graph.

Fig 4 shows the energy consumption graph which compares the energy consumed when KNN algorithm is used against K means algorithm for node size 80. The results show that the KNN algorithm consumes more energy whereas the K means clustering algorithm consumes less energy. 80.

Fig 5 represents the delay graph which shows that the delay produced by the KNN algorithm is more compared to the K means clustering algorithm. The overall delay of KNN algorithm is 40ms whereas the KNN algorithm has a delay of 50ms when the node size maintained is 80. So, the K means is proved to be more efficient method

Fig 6 shows the network overhead for KNN and K means algorithm.The results prove that K means clustering has got less network overhead compared with KNN algorithm when node size is less than 80. When node size increased to 80 and above, the network overhead remains same for both KNN and K means algorithm. So, the Kmeans is suitable for networks with less number of nodes.



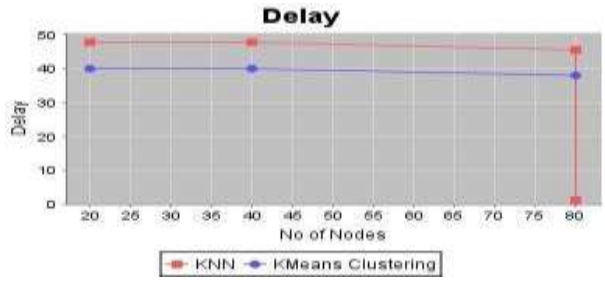
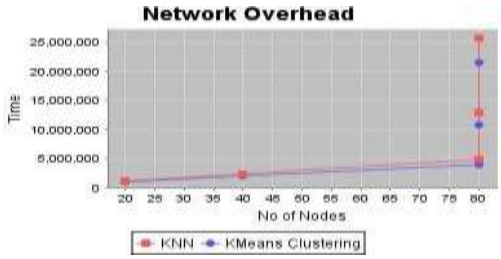


Figure 4: Energy consumption graph **Figure 5: Delay graph**



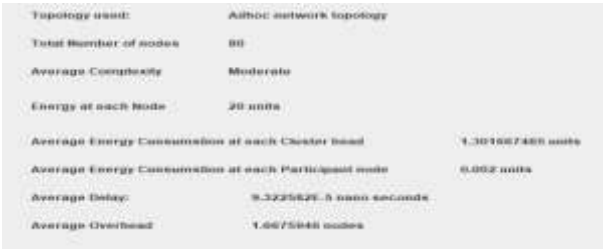


Figure 6: Network overhead graph. **Figure 7: Topology, Complexity, Average Delay,Average**

The simulation results obtained related to network overhead and Energy consumption at each Cluster head ,average delay and average overhead is shown in fig 7.

VI. Conclusion

Wireless Sensor Network applications are gaining popularity these days. Data aggregation is used in Wireless Sensor Networks to save energy by minimizing the number of transmissions. To address the network's constraints, the WSN requires novel solutions. Machine learning algorithms provide a set of approaches for improving a network's ability to adapt to a changing environment. Here used algorithms are K-means and KNN. K-means is a popular clustering algorithm .The larger sensor network data set is reduced to a smaller K-means data set. And the K- method is the most effective for this. While compared with KNN, K-Means as more efficient. Blue nodes in K-Means serve as supplementary nodes, which helps in lowering energy consumption. As a result, we attempted to integrate the greatest aspects of these two approaches. Complexity, latency, overhead, topological awareness, and energy consumption are among the factors used to evaluate algorithms. A progressed similarity-based clustering and information accumulation by utilizing Independent component Analysis is offered to minimize energy utilization in the network.

Future Scope

The introductions of the mobile sink can increase data collection and energy efficiency. In WSNs, numerous issues are still open and require advance investigate efforts.

References

[1] F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E.Cayirci, "Wireless sensor networks: a survey," Computer Networks, 38(4) pp. 393–422, March 2002.

[2] M. A. Alsheikh, S. Lin, D. Niyato and H. P. Tan, Machine Learning in Wireless Sensor Networks: Algorithms, Strategies, and Applications, IEEE Communications Surveys Tutorials, Vol.16, Issue.4, PP.1996- 2018 , 2014.

[3] T. O. Ayodele, "Types of machine learning algorithms," in New Advances in Machine Learning. InTech, 2010.

[4] Hyvärinen, Aapo; Karhunen, Juha; Oja, Erkki (2001). Independent component analysis (Reprint ed.). New York, NY: Wiley. ISBN 978-0-471-40540-5.

[5] S. Siddiqui , A. A. Khan and S. Ghani, A survey on data aggregation mechanisms in wireless sensor network, International Conference of Information and Communication Technologies (ICICT), PP.1-7, 2015

[6] Zaki Ahmad and Abdus Samad, A Study of Machine Learning in Wireless Sensor Network, International Journal of Computer Networks and Applications (IJCNA), Vol.4, Issue. 4, August 2017.

[7] G. Muniraju, S. Zhang, C. Tepedelenlioglu and M. K. Banavar Location Based Distributed Spectral Clustering for Wireless Sensor Networks, IEEE Sensor Signal Processing for Defence Conference (SSPD), 2017.

[8] K. Sohrabi and J. Pottie, "Protocols for self-organization of a wireless sensor network," IEEE Personal Common. 7(5), 2000, pp. 16-27

[9] Jigui Sun, Jie Liu and Lianyu Zhao, "Research on Clustering," Journal of Software, 2008(1):48-49.B. Przydatek, D. Song and A. Perrig, "SIA: Secure Information Aggregation in Sensor Networks," Proc. 1st International Conference on Embedded Networked Sensor Systems (SenSys'03), Los Angeles, CA, Nov. 5-7, 2003, pp. 255–265.

[10] Clustering and Data Aggregation in Wireless Sensor Networks Using Machine Learning Algorithms by Shahina K Research Scholar School of Computing Science and Engineering VIT University, Chennai V. Vaidehi Sr. Professor School of Computing Science and Engineering VIT University, Chennai.

[11] A. Morell, A. Correa and M. Barceló and J. L. Vi cario, Data Aggregation and Principal Component Analysis in WSNs, IEEE Transactions on Wireless Communications, vol.15, Issue.6,PP.3908-3919, 2016.

[12] Data aggregation algorithms for wireless sensor network: A review Mandeep Kaur ∗, Amit Munjal