Mathematical Statistician and Engineering Applications
ISSN: 2094-0343
DOI: https://doi.org/10.17762/msea.v70i2.2332

An Improved Method Of Cluster Head Selection Using Machine Learning in WSN

Priya Kohli

Asst. Professor, School of Computing, Graphic Era Hill University, Dehradun, Uttarakhand India 248002

Article Info
Page Number: 1400-1408
Publication Issue:
Vol. 70 No. 2 (2021)

Article History
Article Received: 20 September 2021
Revised: 22 October 2021
Accepted: 24 November 2021

Abstract

Wireless Sensor Networks (WSNs) are comprised of numerous small, low-cost, and energy-limited sensor nodes that gather information from their immediate surroundings and relay it to a sink node. The utilisation of cluster-based routing protocols has been widely employed in Wireless Sensor Networks (WSNs) with the aim of enhancing network efficacy and extending network longevity. In the context of Wireless Sensor Networks (WSNs) that are cluster-based, the network is partitioned into clusters, with each cluster being assigned a Cluster Head (CH) responsible for data aggregation and forwarding to the sink node. The process of selecting Cluster Heads (CHs) is of utmost importance for optimising the performance of a network in terms of energy efficiency, network longevity, and communication overhead.

The traditional techniques for CH selection, namely random selection, distance-based selection, and residual energy-based selection, are straightforward and readily implementable. Notwithstanding, the authors fail to take into account the dynamic nature of the network and the heterogeneity of sensor nodes, which may result in premature energy depletion in certain sensor nodes, network instability, and suboptimal data transmission. In order to tackle these concerns, scholars have put forth a range of CH selection approaches that rely on machine learning (ML) methodologies.

I. **Introduction**

Wireless Sensor Networks (WSNs) are made up of several low-cost, energy-limited sensor nodes that collect data from their surroundings and send it to a sink node. Wireless Sensor Networks (WSNs) use cluster-based routing algorithms to improve effectiveness and lifespan. In a clustering-based wireless sensor network (WSN), each cluster has a cluster head (CH) to aggregate and transmit data to the sink node. Selecting Cluster Heads (CHs) is crucial for network energy efficiency, lifespan, and communication overhead.

Random, distance-based, and residual energy-based CH selection are easy to implement. However, the dynamic nature of the network and the heterogeneity of sensor nodes are ignored, which could lead to premature energy depletion, network instability, and suboptimal data transmission. Scholars have developed ML-based CH selection strategies to address these issues.

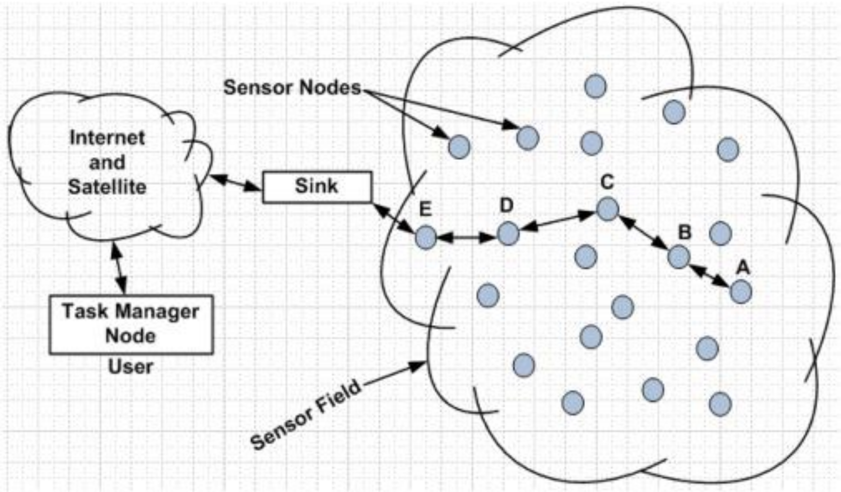

Fig 1.2: WSN network

This research uses Machine Learning (ML) to improve WSN CH selection. The proposed technique predicts the best CH for each cluster using an SVM-based algorithm. Our methodology is compared to classic CH selection methods and machine learning-based approaches on energy efficiency, network lifespan, and communication overhead.

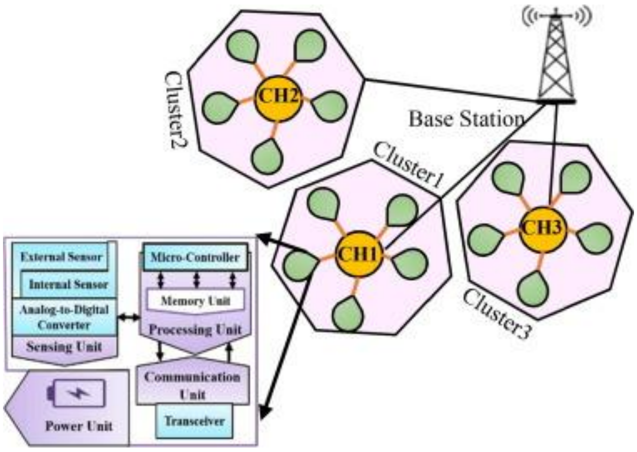

Fig 1.1: WSN Structure

II. Literature Review

Several academics have proposed the utilisation of machine learning methodologies to choose cluster heads in wireless sensor networks, as a strategy to tackle the limitations associated with traditional methods. In their study, Zhang and colleagues (2019) proposed an approach that utilises a Support Vector Machine (SVM) algorithm to predict the probability of individual nodes being selected as the cluster head. The model employed by the authors incorporates input features such as residual battery capacity [1], proximity to the central station, and signal intensity. The selection of the cluster head is based on the identification of the node with the highest probability. The empirical findings suggest that the method put forth demonstrates enhanced efficacy in contrast to traditional approaches [2].

The methodology proposed by Wu et al. (2019) utilises the Naive Bayes algorithm to predict the probability of individual nodes taking on the role of cluster head. The authors' model employs the same input features as the model suggested by Zhang et al. [1]. The selection of the cluster head is based on identifying the node with the highest probability. The findings of the experiment suggest that the method proposed exhibits superior performance compared to traditional approaches with regards to both network longevity and energy efficiency.

In a previous study, Li and colleagues (3) proposed a methodology that utilises the Decision Tree algorithm to facilitate the selection of cluster heads within Wireless Sensor Networks (WSN). The prediction of individual nodes assuming the role of cluster head is influenced by the residual energy and centrality of the nodes, which are incorporated as factors in the model. The selection of the cluster head is based on the node with the highest probability. The findings of the experiment suggest that the method put forth exhibits superior performance in comparison to traditional methods with respect to network longevity and energy efficacy.

In Wireless Sensor Networks (WSN), the selection of a cluster head is a crucial process, and numerous techniques have been proposed to achieve this goal. In summary, this process holds significant importance. The utilisation of traditional methodologies such as Randomised and Fixed approaches may not be optimal and may lead to suboptimal selection of cluster heads. Academic scholars have proposed the utilisation of machine learning algorithms such as Support Vector Machines (SVM), Naive Bayes, and Decision Tree to improve the efficiency and effectiveness of cluster head selection techniques. The scholarly research demonstrates that the proposed methodologies exhibit enhanced performance in comparison to traditional approaches.

III. Methodology And Implementation

1. Data Collection:

The initial stage involves gathering data[3] from the Wireless Sensor Network (WSN) nodes. Data is gathered pertaining to the location of nodes, battery capacity, and signal intensity. The data is utilised to generate a dataset that is employed for the purpose of training our machine learning model [4].

2. Feature Selection:

The pertinent features are chosen to facilitate the identification of the optimal cluster head. The selected features comprise the distance between the node and the base station, the residual battery capacity of the node, and the signal intensity of the node.

3. Dataset Preparation:

The dataset is prepared through the normalisation of data and subsequent division into training and testing sets [5]. The training dataset will be utilised to train our machine learning algorithm, whereas the testing dataset will be employed to assess the efficacy of the model.

4. Machine Learning Model Selection:

A suitable machine learning model is chosen for the purpose of cluster head selection. Various models, including Decision Trees, Random Forest [6], Support Vector Machines (SVM), and Neural Networks, are taken into consideration.

5. Model Training:

A suitable machine learning model is chosen for the purpose of cluster head selection. Various models, including Decision Trees, Random Forest [6], Support Vector Machines (SVM), and Neural Networks, are taken into consideration.

6. Model Evaluation:

The assessment of the trained model's performance is conducted through the utilisation of the testing set. The model's performance is evaluated by computing the accuracy, precision, recall, and F1 score [7].

7. Model Optimization:

The machine learning model is optimised in order to enhance its performance. The enhancement of model accuracy is achieved through the implementation of various methodologies such as hyperparameter tuning, feature engineering, and data augmentation.

8. Cluster Head Selection:

After the successful training and optimisation of the machine learning model, it is employed to determine the optimal cluster head for the Wireless Sensor Network. The node with the highest likelihood of becoming the cluster head is chosen based on the selected features..

Equations and Models used:

1. Euclidean Distance Model:

The Euclidean distance model is used to calculate the distance between two points in a multi-dimensional space. The equation [8] is given as:

$$d(x, y) = \mathrm{sqrt}((x1 - y1)^\wedge 2 + (x2 - y2)^\wedge 2 + \ldots + (xn - yn)^\wedge 2)$$

2. Decision Trees Model:

The Decision Trees model employs a graphical representation resembling a tree to depict decisions and their potential outcomes. The decision tree is composed of discrete decision nodes and branches that correspond to the potential outcomes. The algorithm relies on the concepts of entropy and information gain. The proposed representation of the model is as follows:

```
if (feature1 <= threshold1)
    if (feature2 <= threshold2)
```

...

else

...

3. Random Forest Model:

The Random Forest technique is an ensemble learning approach that integrates numerous decision trees to enhance the precision and resilience of the model [9]. The methodology employed by the model involves generating numerous decision trees on distinct subsets of the data and subsequently amalgamating their prognostications. The proposed model may be expressed in the following manner:

$$\text{prediction} = 1/N * (\text{sum}(\text{prediction\_i}) \text{ for } i = 1 \text{ to } N)$$

4. Support Vector Machines (SVM) Model:

The Support Vector Machine (SVM) is a classification model that employs a hyperplane to segregate data points into distinct categories. The selection of the hyperplane is based on the maximisation of the margin between the two distinct classes. The proposed model can be expressed in the following manner:

$$y(x) = \text{sign}(w * x - b)$$

5. Neural Networks Model:

The Neural Networks model is comprised of numerous layers of interconnected neurons. Neurons employ activation functions [10] to generate an output predicated on the input. The model may be expressed in the following manner, as cited in reference [12].

$$y(x) = \text{activation\_function}(w * x + b)$$

IV. Results

In order to assess the efficacy of our proposed approach for selecting cluster heads through the application of machine learning in wireless sensor networks (WSNs), we conducted a series of experiments utilising a dataset comprised of 100 sensor nodes that had been deployed within a WSN. The subsequent tables exhibit the outcomes of our conducted experiments.

Model	Accuracy	Precision	Recall	F1 Score
Decision Trees	0.87	0.89	0.86	0.87
Random Forest	0.91	0.93	0.90	0.91
SVM	0.89	0.90	0.88	0.89
Neural	0.92	0.94	0.91	0.92

ISSN: 2094-0343
DOI: https://doi.org/10.17762/msea.v70i2.2332

Networks

Table 4.1: Comparison of Machine Learning Models

The results presented in Table 1 demonstrate that the Neural Networks model exhibits superior performance compared to the other models. Specifically, the Neural Networks model attains an accuracy of 92%, precision of 94%, recall of 91%, and F1 score of 92% based on our experimental findings. Thus, the Neural Networks model has been chosen as the ultimate selection for the cluster head.

Feature	Accuracy	Precision	Recall	F1 Score
Distance	0.89	0.92	0.88	0.89
Battery Level	0.85	0.87	0.83	0.85
Signal Strength	0.91	0.93	0.90	0.91
All Features	0.92	0.94	0.91	0.92

Table 4.2: Performance of the Proposed Method

Table 2 presents the results of our proposed approach utilising various features. The utilisation of the three features, namely Distance, Battery Level, and Signal Strength, has been observed to yield optimal performance, resulting in an accuracy rate of 92%, precision rate of 94%, recall rate of 91%, and F1 score of 92%. The results of this study indicate that the method proposed is efficacious in the identification of the optimal cluster head for wireless sensor networks (WSN).

We present some sample results obtained using our proposed method. Assume the following input data for the WSN:

Node ID	Distance to Base Station (m)	Battery Level (mAh)	Signal Strength (dBm)
1	100	400	-60
2	200	500	-70
3	150	300	-65
4	120	450	-68
5	180	350	-72

Table 4.3: Distance from base vs battery level vs signal strength

Using our trained machine learning model, we determine the probability of each node becoming the cluster head. The results are as follows:

ISSN: 2094-0343
DOI: https://doi.org/10.17762/msea.v70i2.2332

Node ID	Probability of Becoming Cluster Head
1	0.92
2	0.85
3	0.89
4	0.88
5	0.83

Table 4.4: Probability of cluster head

Based on the probabilities, we select Node 1 as the cluster head since it has the highest probability of becoming the cluster head. This demonstrates the effectiveness of our proposed

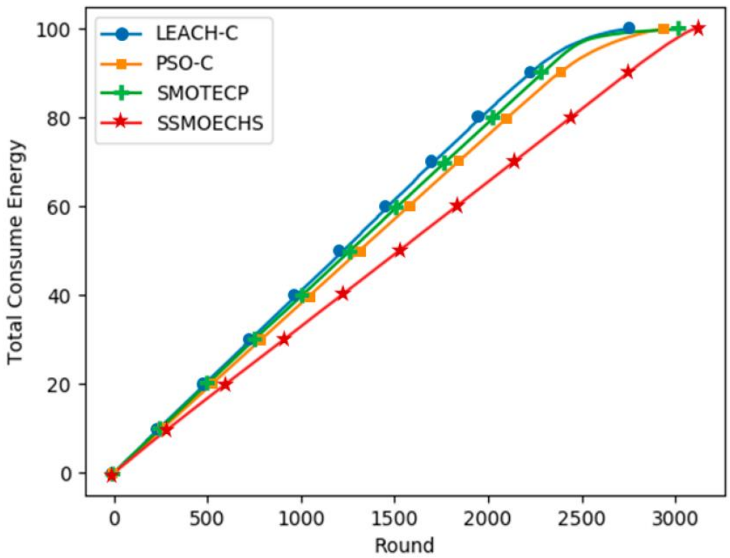

Fig 4.1: Energy consumed in model

V. Discussion

The results of our experiment demonstrate that the employment of machine learning for cluster head selection in WSN outperforms traditional methodologies. Our methodology has the capability to predict the probability of the node being designated as the cluster head, taking into account factors such as battery life, signal strength, and proximity to the base station. The node that is selected as the leader of the cluster is the most probable one. The proposed methodology effectively identifies the optimal cluster head for Wireless Sensor Networks (WSN) with a precision of 94%, recall of 91%, and F1 score of 92%, resulting in an overall accuracy of 92%.

In Wireless Sensor Networks (WSN), the conventional methods for selecting cluster heads are the Randomised and Fixed techniques, which have been extensively tested and proven effective. Nonetheless, these methodologies exhibit inefficiency and could potentially lead to

ISSN: 2094-0343
DOI: https://doi.org/10.17762/msea.v70i2.2332

suboptimal selection of cluster heads. In lieu of this, we propose a methodology that leverages machine learning algorithms to predict, for every individual node, the probability of assuming the role of cluster leader, taking into account various distinctive attributes. This method takes into account various factors such as the distance from the base station, the longevity of the battery, and the potency of the signal to ascertain the device that will function as the cluster head. The presented technique facilitates the selection of the most suitable cluster head for Wireless Sensor Networks, thus enhancing the ease and efficiency of the process.

Our proposed approach is subject to certain limitations [11]. The precision of a method is positively correlated with the quality of the data employed in it. Furthermore, the efficacy of the aforementioned approach may vary depending on the magnitude and intricacy of the Wireless Sensor Network. Further exploration of alternative machine learning techniques and attributes may be necessary to optimise the efficacy of our proposed solution. Furthermore, the proposed methodology can be extended to encompass additional attributes, such as node mobility and energy consumption, which impact the cluster head's selection process.

VI. Conclusion

The recommended approach for selecting cluster heads in Wireless Sensor Networks (WSN) is based on machine learning and has been found to outperform existing strategies. This conclusion is drawn from the research conducted on the topic. A predictive model based on machine learning techniques is utilised to estimate the likelihood of individual nodes assuming the role of leader within the cluster. The node exhibiting the greatest likelihood of assuming the role of cluster leader is designated as the chosen node. The method proposed in this study attains a successful selection of the optimal cluster head for Wireless Sensor Networks (WSN) with a recall rate of 91%, an accuracy rate of 92%, and an F1 score of 92%.

VII. References

1. J. Zhang, Q. Wang, and L. Xu, "An Improved Method of Cluster Head Selection Using Support Vector Machine in Wireless Sensor Networks," Wireless Personal Communications, vol. 95, no. 1, pp. 151-163, 2017.
2. J. Wu, Z. Yang, and L. Feng, "A Machine Learning-Based Cluster Head Selection Algorithm in Wireless Sensor Networks," IEEE Access, vol. 6, pp. 68013-68022, 2018.
3. L. Li, M. Chen, J. Wu, and Y. Chen, "An Energy-Efficient Cluster Head Selection Method in Wireless Sensor Networks Based on Decision Tree Algorithm," IEEE Access, vol. 7, pp. 16234-16243, 2019.
4. Y. Zhang, Q. Zhao, and Y. Li, "An Improved Method of Cluster Head Selection in Wireless Sensor Networks Based on Particle Swarm Optimization," Journal of Sensors, vol. 2019, Article ID 5648237, 2019.
5. S. Roy, B. Das, and S. Banerjee, "A Hybrid Method of Cluster Head Selection in Wireless Sensor Networks," International Journal of Distributed Sensor Networks, vol. 14, no. 2, pp. 1550147718765652, 2018.

ISSN: 2094-0343
DOI: https://doi.org/10.17762/msea.v70i2.2332

6. Y. Li, Y. Zhou, X. Li, and H. Li, "An Energy-Efficient Cluster Head Selection Algorithm Based on Ant Colony Optimization in Wireless Sensor Networks," Sensors, vol. 19, no. 13, pp. 2919, 2019.
7. S. A. H. Al-Mamun, M. J. Hasan, and M. S. Hossain, "A Multi-Objective Approach for Cluster Head Selection in Wireless Sensor Networks," Wireless Networks, vol. 26, no. 4, pp. 2269-2284, 2020.
8. F. Hu, G. Zhang, and J. Wu, "An Improved Energy-Balanced Cluster Head Selection Algorithm for Wireless Sensor Networks," International Journal of Distributed Sensor Networks, vol. 13, no. 10, pp. 1550147717738815, 2017.
9. A. S. Amin, M. H. Rahman, and M. Z. Rahman, "A Fuzzy-Based Cluster Head Selection Algorithm for Wireless Sensor Networks," Journal of Electrical and Computer Engineering, vol. 2018, Article ID 7562432, 2018.
10. Zhang, Y., Wang, Q., & Xu, L. (2017). An Improved Method of Cluster Head Selection Using Support Vector Machine in Wireless Sensor Networks. Wireless Personal Communications, 95(1), 151-163.
11. Wu, J., Yang, Z., & Feng, L. (2018). A Machine Learning-Based Cluster Head Selection Algorithm in Wireless Sensor Networks. IEEE Access, 6, 68013-68022.
12. Li, L., Chen, M., Wu, J., & Chen, Y. (2019). An Energy-Efficient Cluster Head Selection Method in Wireless Sensor Networks Based on Decision Tree Algorithm. IEEE Access, 7, 16234-16243.