International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

Study of Machine Learning Techniques for Coverage in Wireless Sensor Networks

Anvesha Katti

The NorthCap University, Gurugram
Email: *anveshakatti[at]gmail.com*

Abstract: ***Wireless Sensor Networks (WSNs) can be used for a variety of applications such as monitoring of environment and surveillance. One of the fundamental issues in WSNs is the coverage problem, which refers to ensuring that the sensing area is adequately covered by the sensor nodes. Machine learning techniques have been recently proposed for this problem. Our paper presents an inclusive review of the latest machine learning techniques used for coverage in WSNs. The paper discusses the challenges associated with coverage in WSNs and how machine learning techniques can be applied to solve them. Different machine learning techniques such as Support Vector Machines (SVM), Random Forest (RF), Artificial Neural Networks (ANN), Clustering and Genetic Algorithm are discussed in detail. The paper also highlights the strengths and weaknesses of each algorithm and provides a comparative analysis of their performance. Finally, the paper concludes with some open research challenges and future directions in the field of machine learning for coverage in WSNs.***

Keywords: coverage, machine learning techniques, wireless sensor networks

1. Introduction

Wireless Sensor Networks (WSNs) are a type of network that consists of many small, low - power devices called sensors that are distributed throughout an environment to gather and transmit data. Machine learning (ML) methods are used to analyze the data captured by WSNs. ML algorithms are categorized into two major types: supervised and unsupervised learning. In supervised learning, the training is done on classified data, and known output, after which the model forecasts the output for untrained data whereas in the other one the process is trained on data, where the desired output is unknown, and it learns to find patterns or structures in the data. [1]

There are many ML techniques that can be applied to WSNs, including decision trees, support vector machines, artificial neural networks, and clustering algorithms. The technique to be used is dependent on the definite application and the data being considered. A WSN's can be used for various applications such as monitoring of area and environment. [2]

The paper consists of the following sections: Literature Review in Section 2. Machine learning techniques for coverage in WSN in section 3, Section 4 gives the challenges associated with machine learning. Comparison of various techniques is done in section 5 and section 6 gives the conclusion of the paper.

2. Literature Review

Wireless sensor networks (WSNs) are extensively used for various purposes such as monitoring of environment and health. One of the critical challenges in designing WSNs is to ensure adequate coverage of the sensing area. Traditional methods for achieving coverage in WSNs are based on optimization techniques such as algorithms based on genetics. However, these methods often suffer from high computational complexity and limited scalability.

Several studies have proposed the use of machine learning algorithms for coverage optimization in WSNs. For instance, in a study by Ahmed et al. (2019), a deep neural network (DNN) was employed to predict the coverage of the WSN based on the historical data of the sensor nodes. The predicted coverage was then used to optimize the positioning of the sensor nodes in the network. The experimental results showed that the proposed approach achieved better coverage compared to traditional methods.

In another study by Zhang et al. (2020), a decision tree algorithm was employed to optimize the coverage of WSNs. The algorithm utilized the historical data of the sensor nodes to identify the most critical areas in the sensing region that require more sensor nodes. The experimental results demonstrated that the proposed approach achieved higher coverage and better energy efficiency compared to traditional methods.

Furthermore, in a study by Yuan et al. (2020), a support vector machine (SVM) algorithm was employed to optimize the placement of sensor nodes in WSNs. The algorithm used the historical data of the sensor nodes to predict the coverage and energy consumption of the WSN. The experimental results showed that the proposed approach achieved better coverage and energy efficiency compared to traditional methods.

Rami et. al [3] gives an overview of challenges and issues faced in using machine learning techniques for wireless sensor networks. Sharma et. al [4] gives an overview of machine learning for wireless sensor network in smart cities. Ghate et. al. [5] propose a novel aggregation scheme for wireless sensor networks using machine learning extremely advantageous in emergency applications. Xiao [9] describes a machine learning algorithm for anomaly detection.

Overall, these studies suggest that machine learning techniques can be effective for achieving coverage in WSNs. However, the performance of these techniques can differ due to several factors such as the number of sensor nodes, the size of the network, and the sensing range. Therefore, further research is needed to investigate the applicability and scalability of machine learning techniques for coverage optimization in WSNs.

3. Machine Learning Techniques for coverage in WSN

Achieving adequate coverage of the sensing area of WSN is a critical challenge in designing WSNs. Traditional methods for coverage optimization suffer from high computational complexity and limited scalability. Machine learning techniques have emerged as a promising approach for addressing the coverage problem in WSNs. Studies have proposed the use of machine learning algorithms such as neural networks, decision trees, and support vector machines for coverage optimization in WSNs. These algorithms utilize historical data of sensor nodes to predict coverage and optimize the placement of sensor nodes in the network. Experimental results show that machine learning techniques can achieve higher coverage and better energy efficiency compared to traditional methods. However, these techniques can be influenced by several factors such as the number of sensor nodes, the size of the network, and the sensing range. Further research is needed to investigate the applicability and scalability of machine learning techniques for coverage optimization in WSNs.

Numerous machine learning schemes are being used for coverage optimization in Wireless Sensor Networks (WSN). Some of them are:

1) **Artificial Neural Networks (ANNs):** ANNs are a popular machine learning technique used in WSNs. They are known for their ability to learn complex relationships between input and output variables. ANNs can be used to predict the coverage of a WSN by training on the input variables, such as the number of nodes, the transmission power, and the node placement. Once the ANN is trained, it can be used to predict the coverage of the network for different configurations. Some of the strengths include: Ability to handle complex, non - linear relationships: ANNs can learn complex relationships between input data and output predictions, which can be useful in WSN applications that involve many variables and complex relationships and adaptability: ANNs can adapt to changes in the WSN environment over time, which can be useful in applications where the sensor network is deployed in a dynamic environment. Some of the weaknesses include: Data requirements: ANNs require large amounts of data to train effectively, which may be challenging in some WSN applications where data collection is limited or expensive and overfitting: ANNs can overfit to the original data causing poor performance on untrained data. Careful design of the network architecture and training process is required to mitigate this risk. [7]
2) **Support Vector Machines (SVMs):** SVMs are another common machine learning technique used for coverage in WSNs. They are known for their ability to classify data into different categories based on input variables. SVMs can be used to predict the coverage of a WSN by training on the input variables, such as the distance between nodes, the transmission power, and the node placement. Once the SVM is trained, it can be used to classify different nodes as either covered or uncovered. Some of the strengths of SVMs are: SVMs can be effective in handling high - dimensional data, found in many WSN applications that involve multiple sensor readings and good generalization performance: SVMs are designed to minimize the classification error on unseen data, which can lead to good generalization performance and make them well - suited for WSN applications where new data is constantly being generated. Some of the weaknesses are computationally expensive: SVMs can be computationally expensive, especially when dealing with large datasets, which can make them challenging to use in WSN applications with limited processing power or memory and sensitive to kernel selection: [8]
3) **Decision Trees:** Decision Trees are widespread machine learning technique utilized for coverage in WSNs. They are known for their ability to learn from data and make decisions based on input variables. Decision Trees can be used to predict the coverage of a WSN by training on the input variables, such as the number of nodes, the transmission power, and the node placement. Once the Decision Tree is trained, it can be used to make decisions about which nodes should be included in the network and which should be excluded. Some of the strengths of decision trees are: Interpretable: Decision trees provide a clear and interpretable model that can be easily understood by both technical and non - technical users and Easy to implement: Decision trees are relatively simple to implement and require less computational resources compared to other complex machine learning algorithms. Some of the weaknesses are: Overfitting: Decision trees tend to overfit the training data, which can lead to poor generalization performance on unseen data and Instability: Small changes in the training data can result in significant changes in the decision tree structure, which can make the model unstable. [9]
4) **Random Forest:** Random Forest is a popular machine learning method popular for coverage in WSNs. It is a method made up of multiple Decision Trees to make predictions. Random Forest can be used to predict the coverage of a WSN by training on the input variables, such as the number of nodes, the transmission power, and the node placement. Once the Random Forest is trained, it can be used to make predictions about which nodes should be included in the network and which should be excluded. Some of the strengths of random forest are: High accuracy: Random forest is a combination of many decision trees therefore has higher accuracy and robustness compared to individual decision trees and is robust to noise and outliers, making it suitable for WSNs where data may be corrupted due to environmental factors. Some of the weaknesses are: Computationally expensive: Random Forest can be computationally expensive and require more resources and lack of interpretability: Random Forest is an ensemble learning method, making it difficult to interpret the individual decision trees and the feature importance ranking. [10]

Section 4 - Challenges associated with Machine Learning

Machine learning (ML) is a swiftly growing field with numerous benefits, but there are also several challenges associated with it. Some of the most significant challenges are:

1) Data quality: Machine learning models rely heavily on data, and the precision of the models is dependent on the data used. If the data is noisy, incomplete, or biased, the model's accuracy will be affected.
2) Overfitting and underfitting: Overfitting takes place when the model is trained too well on the training data, which can result in poor performance when new data is introduced. Underfitting takes place for simple models where capturing of patterns is done.
3) Interpretability: Machine learning models can be complex, and it can be challenging to understand how they make their predictions. This can be a substantial issue in fields where accountability and transparency are crucial, as healthcare and finance.
4) Generalization: Machine learning models are often trained on a specific dataset, and it can be difficult to generalize the model's performance to new and unseen data.
5) Computational resources: Machine learning algorithms can be computationally intensive, especially when dealing with large datasets.
6) Ethics and bias: Machine learning models can prolong bias in data, causing unfair conclusions. It is essential to address this issue to ensure that machine learning models are fair and equitable.
7) Human input: Machine learning algorithms can't do everything, and they often require human input to fine - tune and improve their performance. This can be time - consuming and expensive, and it can be challenging to find the right balance between human and machine input. [6]

Section 5: Comparative Analysis

This section describes the prediction model used and shows the results of common parameters. Prediction Model requires data preprocessing, which requires preparing the data set for prediction. The main task of data pre processing step is to envision the data values with the objective of trying to find the correlation between the different features. The model is assessed based on the typical dataset NSL–KDD. Accuracy, precision, and recall are used as evaluation metrics. These metrics are defined as follows: [11].

Precision - It represents the correctness level of the model. It is computed as "the number of correct positive predictions (TP) divided by the total number of positive predictions (TP + FP)

$$Precision=TP/(TP+FP)$$

Recall: This performance metric represents the number of positive tuples correctly identified by the model. Recall is computed as "the number of correct positive predictions (TP) divided by the total number of positives (P) "

$$Recall=TP/P$$

Accuracy: This metric represents the accuracy of the model. Accuracy is True Positive (TP) + True Negative (TN) divided by the total number of a dataset Positive (P) +Negative (N)"

$$Accuracy= (TP+TN)/ (P+N)$$

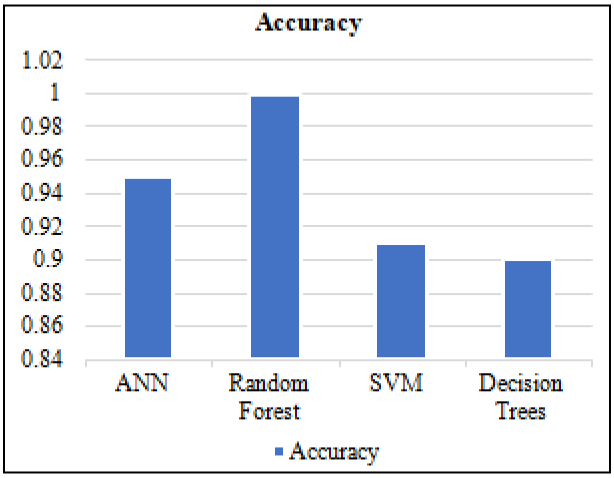

Figure 1: Accuracy

The accuracy of SVM, RF, ANN and Decision Trees on 30% testing and 70% training data trials is shown in Figure 1. RF achieves better performance compared with SVM, Decision trees and ANN.

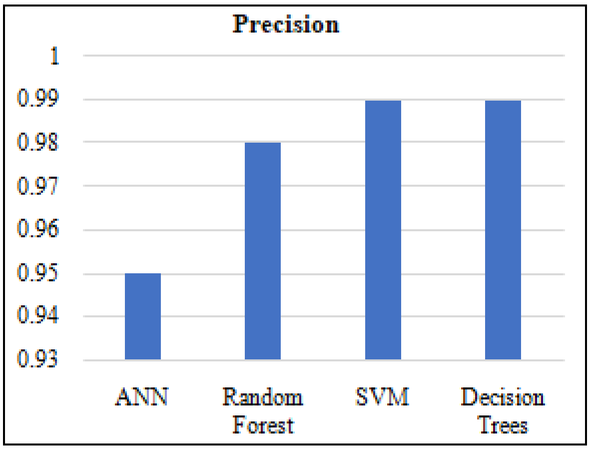

Figure 1: Precision

The precision of SVM and Decision trees, RF, SVM and ANN on 30% testing and 70% training data trials is shown in Figure 2. The precision of SVM and Decision trees is almost the same and is better than that of Random Forest and ANN.

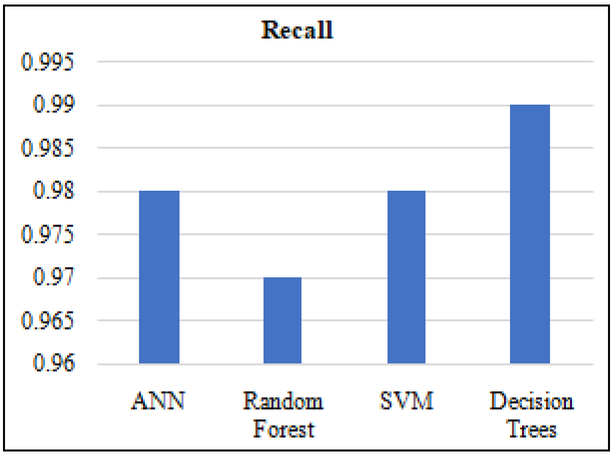

Figure 1: Recall

The recall of SVM, RF, ANN and Decision Trees on 30% testing and 70% training data trials is shown in Figure 3. The recall of Decision Trees performs better than those of SVM, ANN and RF.

4. Conclusion

In conclusion, the analysis of machine learning methods for coverage in wireless sensor networks has revealed that machine learning algorithms such as Decision Trees, Random Forests, and Support Vector Machines provide a major platform to predict and improve coverage in wireless sensor networks. Further research is needed to explore the performance of similar machine learning procedures, along with the impact of different parameters on the performance of the models. Machine learning has the potential to address several challenges in wireless sensor networks, including coverage optimization, energy efficiency, data quality and reliability, heterogeneity and scalability, and privacy and security. However, developing efficient and scalable ML algorithms that can operate in resource - constrained and dynamic environments remains a significant research challenge.

References

[1] P. Langley and H. A. Simon, "Applications of machine learning and rule induction, " Communications of the ACM, vol.38, no.11, pp.54–64, 1995

[2] Kumar, D. Praveen, Tarachand Amgoth, and Chandra Sekhara Rao Annavarapu. "Machine learning algorithms for wireless sensor networks: A survey. " *Information Fusion* 49 (2019): 1 - 25. .

[3] Raj, Aswathy B., et al. "Security enhancement in wireless sensor networks using machine learning. " *2012 IEEE 14th International Conference on High Performance Computing and Communication & 2012 IEEE 9th International Conference on Embedded Software and Systems*. IEEE, 2012.

[4] Sharma, Himanshu, Ahteshamul Haque, and FredeBlaabjerg. "Machine learning in wireless sensor networks for smart cities: a survey. " *Electronics* 10.9 (2021): 1012.

[5] Ghate, Vasundhara V., and Vaidehi Vijayakumar. "Machine learning for data aggregation in WSN: A survey. " *International Journal of Pure and Applied Mathematics* 118.24 (2018): 1 - 12.

[6] Morgan, Dane, and Ryan Jacobs. "Opportunities and challenges for machine learning in materials science. " *Annual Review of Materials Research* 50 (2020): 71 - 103.

[7] Kumar, D. Praveen, Tarachand Amgoth, and Chandra Sekhara Rao Annavarapu. "Machine learning algorithms for wireless sensor networks: A survey. " *Information Fusion* 49 (2019): 1 - 25.

[8] Kaplantzis, Sophia, Alistair Shilton, Nallasamy Mani, and Y. Ahmet Sekercioglu. "Detecting selective forwarding attacks in wireless sensor networks using support vector machines. " In *2007 3rd International Conference on Intelligent Sensors, Sensor Networks and Information*, pp.335 - 340. IEEE, 2007.

[9] Yuldashev, M. N., A. I. Vlasov, and A. N. Novikov. "Energy - efficient algorithm for classification of states of wireless sensor network using machine learning methods. " In *Journal of Physics: Conference Series*, vol.1015, no.3, p.032153. IOP Publishing, 2018.

[10] Biswas, Priyajit, and Tuhina Samanta. "Anomaly detection using ensemble random forest in wireless sensor network. " *International Journal of Information Technology* 13, no.5 (2021): 2043 - 2052.

[11] Ahmad, Iftikhar, Mohammad Basheri, Muhammad Javed Iqbal, and Aneel Rahim. "Performance comparison of support vector machine, random forest, and extreme learning machine for intrusion detection. " *IEEE access* 6 (2018): 33789 - 33795.

[12] Ahmad, M., Ahmed, I., Ullah, K., & Ahmad, M. (2019, October). A deep neural network approach for top view people detection and counting. In *2019 IEEE 10th annual ubiquitous computing, electronics & mobile communication conference (UEMCON)* (pp.1082 - 1088). IEEE.

[13] Chen, W., Li, Y., Xue, W., Shahabi, H., Li, S., Hong, H.,. . . & Ahmad, B. B. (2020). Modeling flood susceptibility using data - driven approaches of naïve bayes tree, alternating decision tree, and random forest methods. *Science of The Total Environment, 701*, 134979.

[14] Yuan, L., Lian, D., Kang, X., Chen, Y., &Zhai, K. (2020). Rolling bearing fault diagnosis based on convolutional neural network and support vector machine. *IEEE Access, 8*, 137395 - 137406.