JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

MACHINE LEARNING TECHNIQUES FOR WIRELESS SENSOR NETWORKS

[1]Ashwini G. Dhumal, [2]Mairaj U. Inamdar
[1]PG Student, [2]Assistant Professor
Department of Electronics and Telecommunication Engineering
Siddhant College of Engineering, Pune, India

Abstract: Wireless Sensor Networks (WSNs) play a crucial role in various domains, including environmental monitoring, healthcare, surveillance, and industrial automation. With the increasing complexity and scale of WSNs, there is a growing need for efficient data processing and decision-making mechanisms. Machine learning techniques have emerged as powerful tools to leverage the vast amount of data collected by WSNs and extract valuable insights. This research paper provides a comprehensive review and analysis of machine learning techniques applied to WSNs. The paper begins by introducing the fundamental concepts of WSNs, including their architecture, and energy constraints. It then delves into the potential applications and challenges faced by WSNs, such as limited resources, data heterogeneity, and dynamic environments, security. A detailed survey of machine learning techniques suitable for WSNs is presented, including supervised learning, unsupervised learning, and reinforcement learning. Various algorithms within these categories, such as decision trees, support vector machines, clustering methods, and deep learning models, are discussed in terms of their applicability and performance in WSNs. It explores how machine learning algorithms can enhance network performance, optimize energy consumption, and improve data analysis capabilities in WSNs.

***Index Terms* - Wireless sensor networks, Machine learning, security**

I. INTRODUCTION

A wireless sensor network (WSN) often consists of a large number of independent, low-cost, small, low-power sensing devices having sensing capabilities. These sensors collect data from their surroundings and work together to send the collected data to a central sink or base station for further processing. Sensor nodes can accommodate a variety of sensors including thermal, acoustic, chemical, pressure, weather, and optical sensors. Because of these differences, WSNs have great potential for powerful applications, each with its own characteristics and requirements [1].

A wireless sensor network (WSN) entails distributed tiny sensor nodes which might be supposed to screen bodily or environmental conditions and communicate with every different and alternate facts and information. Due to their small size, sensor nodes have limited computational power and energy resources. Additionally, the environment wherein they are placed varies dramatically over time. It is essentially critical to analyze sensor data as soon as it is collected. Sensor node data that has not been processed for a long duration is assumed as incomplete and inaccurate. Since WSNs are usually dynamic in nature, their topologies will frequently change. As a result of the connection loss, the network needs to add a new node. The future scope of WSN technology is bright across a wide range of application areas. This paper lists a few of the most useful ones and also how the different machine learning (ML) techniques are used in deploying the various sensor networks. However there are various other issues when it comes to these networks.

Wireless Sensor Networks (WSNs) face several challenges that need to be addressed for their effective deployment and operation. Some of the key challenges include:

1) Limited Energy Resources: Sensor nodes in WSNs are typically battery-powered and have limited energy resources. Energy efficiency is critical to prolong the network's operational lifetime. Optimizing energy consumption at various levels, such as data transmission, processing, and sensing, is a major challenge.
2) Scalability: WSNs often comprise a large number of sensor nodes deployed over a wide area. As the network size grows, scalability becomes a challenge. Efficient management of a large number of nodes, data aggregation, and routing algorithms that can handle network scalability is essential.
3) Communication Constraints: Wireless communication in WSNs is subject to constraints such as limited bandwidth, limited transmission range, and potential interference from other devices or environmental factors. These constraints must be considered when designing communication protocols and data transmission strategies.
4) Data Management: WSNs generate enormous amounts of data from multiple sensors. Efficient data management, including data acquisition, storage, processing, and retrieval, is crucial. Data compression, aggregation, and fusion techniques are employed to reduce the amount of data transmitted and stored.
5) Node Localization: Accurate localization of sensor nodes is essential for many applications, such as target tracking, environmental monitoring, and surveillance. However, achieving precise localization in WSNs can be challenging due to factors like signal attenuation, environmental obstacles, and the need for localization algorithms that are energy-efficient.

6) Security and Privacy: WSNs are vulnerable to various security threats, including unauthorized access, tampering, and data interception. Ensuring secure communication, data integrity, and privacy protection are critical challenges in WSNs. Resource-constrained sensor nodes require lightweight and efficient security mechanisms.
7) Reliability and Fault Tolerance: Sensor nodes can be prone to failures due to factors such as environmental conditions, hardware faults, or battery depletion. Designing fault-tolerant mechanisms and ensuring reliable operation in the presence of node failures or network disruptions is a significant challenge.

In the late 1950s, artificial intelligence (AI) techniques such as machine learning (ML) were first presented [2]. As time went on, the emphasis changed to more resilient and computationally feasible techniques. Machine learning methods have been widely applied in the past ten years for a wide range of tasks, such as classification, regression, and density estimation, in a number of application areas, including computer vision, bioinformatics, speech recognition, spam detection, fraud detection, and advertising networks. Numerous interdisciplinary topics, including as statistics, mathematics, neuroscience, and computer science, are represented in the methods and approaches employed. Machine learning is important in WSN applications for the following main reasons:

1) Big data issues arise from WSNs' huge data generation from numerous sensors. Based on the patterns and relationships found in the data, ML algorithms are able to process and analyze this data efficiently, extract insightful information, and make defensible conclusions.
2) Because sensor nodes have limited energy resources, energy efficiency is a major challenge in WSNs. By cleverly controlling data transfer, processing, and sensing tasks, machine learning algorithms can maximize energy use. Techniques for energy optimization based on machine learning (ML) can increase overall energy efficiency and prolong the network's operating lifetime.
3) WSNs operate in dynamic and changing environments. ML techniques enable WSNs to adapt to varying conditions by continuously learning and updating their knowledge based on new data. ML algorithms can adjust network behavior, optimize resource allocation, and improve performance in response to changing environmental factors.
4) ML algorithms can assist in fault detection and localization in WSNs. By analyzing sensor data patterns, ML techniques can identify abnormal behavior or faults in the network. ML-based fault detection systems can help in timely identification and mitigation of faults, ensuring the reliability and robustness of the network.
5) WSNs are vulnerable to security threats, and ML techniques can enhance their security by detecting and mitigating these threats. ML algorithms can analyze network traffic patterns, identify anomalies, and detect intrusion attempts. ML-based intrusion detection systems improve the security of WSNs by providing early warning and preventing unauthorized access or malicious activities.

The paper is organized as, Section II describes a literature review of existing machine learning techniques to address issues in WSNs .Complete description of Machine Learning Techniques for Wireless Sensor Networks is given in Section III. Section IV gives conclusion of the paper.

II. LITERATURE REVIEW

Wireless Sensor Network many challenges related to security due to its dynamic nature. In paper [3], an attempt is made to investigate the security related issues, the challenges and to propose some solutions to secure the WSN against these security threats. While the set of challenges in sensor networks are diverse, this paper focus only on the challenges related to the security of Wireless Sensor Network also summarizes the attacks and their classifications in wireless sensor networks. The author of [4], describes several works that applied machine learning techniques and deep learning techniques on diverse research areas including networking, communications and lossy environment. Further the survey identify the possible issues and challenging tasks for applying the different deep learning and machine learning algorithms and strategies in wireless networks. The paper [5] overviews common issues in wireless sensor networks (WSNs) and how they can be resolved using several machine learning methods.

In [6], a short survey of machine learning algorithms applied in WSNs for information processing and for improving network performance was presented. A related survey that discussed the applications of machine learning in wireless ad-hoc networks was published in [7]. The authors of [8] discussed applications of three popular machine learning algorithms (i.e., reinforcement learning, neural networks and decision trees) at all communication layers in the WSNs. In contrast, specialized surveys that touch on machine learning usage in specific WSN challenges have also been written. For instance, [9], [10] addressed the development of efficient outlier detection techniques so that proper actions can be taken, and some of these techniques are based on concepts from machine learning. Meanwhile, [11] discusses computational intelligence methods for tackling challenges in WSNs such as data aggregation and fusion, routing, task scheduling, optimal deployment and localization. Here, computational intelligence is a branch of machine learning that focuses on biologically inspired approaches such as neural networks, fuzzy systems and evolutionary algorithms [12].Several important task occur in WSN which have to perform priority-wise. This task can be further optimized with the help of ML approaches [13]. The sensor node has limited energy and bandwidth constraint, routing overhead is created in the transmission of the data from sender to receiver. ML approach provides the mechanisms which minimize the energy requirement and provide new energy management scheme for reliable communication between sender and receiver [14].

Generally, these early surveys concentrated on reinforcement learning, neural networks and decision trees which were popular due to their efficiency in both theory and practice. In this paper, we decided instead to include a wide variety of important up-to-date machine learning algorithms for a comparison of their strengths and weaknesses. In particular, we provide a comprehensive overview which groups these recent techniques roughly into supervised, unsupervised and reinforcement learning methods.

III. MACHINE LEARNING TECHNIQUES FOR WIRELESS SENSOR NETWORKS

Machine learning is typically defined by sensor network designers as a set of tools and techniques used to build prediction models. Machine learning experts acknowledge that it is a rich topic with numerous, expansive themes and patterns. Those who want to apply machine learning to WSNs will find it helpful to comprehend such themes. Machine learning algorithms offer great flexibility benefits when applied to various WSNs applications. In the context of WSNs, this section offers some theoretical ideas and approaches for implementing machine learning.

The proposed structure of the model can be used to categorize existing machine learning techniques. The majority of machine learning algorithms can be classified as either supervised or unsupervised as well as reinforcement learning [15]. A labeled training data set is given to machine learning algorithms in the first category. Using this set, the system model that depicts the discovered relationship between the input, output, and system parameters is constructed. Unsupervised learning methods, in contrast to supervised learning, do not receive labels (i.e., an output vector). Basically, the goal of an unsupervised learning algorithm is to classify the sample sets to different groups (i.e., clusters) by investigating the similarity between the input samples. The third category includes reinforcement learning algorithms, in which the agent learns by interacting with its environment (i.e., online learning).

A. Supervised Learning

It is an approach that plays a very important role in ML. In this type of learning, the machine offers labelled input and expected (desired) output. The main aim of this type of learning approach is that an algorithm developed is capable of learning automatically by comparing the actual output with the expected output .This difference between the actual and expected output is known as an error. Supervised learning algorithms are extensively used to solve several challenges in WSNs such as localization and objects targeting, event detection and query processing, media access control, security and intrusion detection and quality of service (QoS), data integrity and fault detection.

Supervised learning can be classified into two categories: Regression and Classification. Regression is the simplest ML approach which provide accurate and precise outcomes. In this approach, the value of Y attribute is described which is dependent on the value of the X attribute. This model is quantitative and continuous in nature. Classification is a technique used to predict discrete values and include approaches such as Decision Tree, Artificial Neural Network (ANN), Random Forest (RF), Support Vector Machine (SVM), Bayesian and k-Nearest Neighbor (k-NN). Some of the supervised learning techniques are explained below:

1) K-nearest neighbor (k-NN): This supervised learning algorithm classifies a data sample (called a query point) based on the labels (i.e., the output values) of the near data samples. For example, missing readings of a sensor node can be predicted using the average measurements of neighboring sensors within specific diameter limits. There are several functions to determine the nearest set of nodes. A simple method is to use the Euclidean distance between different sensors. K-nearest neighbor does not need high computational power, as the function is computed relative to local points (i.e., k-nearest points, where k is a small positive integer). This factor coupled with the correlated readings of neighboring nodes makes k-nearest neighbor a suitable distributed learning algorithm for WSNs. In [16], it has been shown that the k-NN algorithm may provide inaccurate results when analyzing problems with high-dimensional spaces (more than 10-15 dimensions) as the distance to different data samples becomes invariant (i.e., the distances to the nearest and farthest neighbors are slightly similar). In WSNs, the most important application of the k-nearest neighbor algorithm is in the query processing subsystem.

2) Decision tree (DT): It is a classification method for predicting labels of data by iterating the input data through a learning tree [17]. During this process, the feature properties are compared relative to decision conditions to reach a specific category. The literature is very rich with solutions that use DT algorithm to resolve different WSNs' design challenges. For example, DT provides a simple, but efficient method to identify link reliability in WSNs by identifying a few critical features such as loss rate, corruption rate, mean time to failure (MTTF) and mean time to restore (MTTR). However, DT works only with linearly separable data and the process of building optimal learning trees is NP-complete [18].

3) Support vector machines (SVMs): Support Vector Machines offer alternatives for neural networks that are preferred options for solving non convex unconstrained optimization problems. In the context of WSN, they have been used for intrusion detection or detecting the malicious behavior of sensor nodes, security, and localization. With SVM, it is possible to uncover the spatio-temporal correlations in data, as the algorithm involves constructing a set of hyper planes separating WSN data measurements in feature space, by as wide as possible margins.

4) Neural networks (NNs): This learning algorithm could be constructed by cascading chains of decision units used to recognize nonlinear and complex functions [19]. In WSNs, using neural networks in distributed manners is still not so pervasive due to the high computational requirements for learning the network weights, as well as the high management overhead. However, in centralized solutions, neural networks can learn multiple outputs and decision boundaries at once, which makes them suitable for solving several network challenges using the same model.

5) Naïve Bayes: The naïve Bayes is a classification method based on Bayes' theorem and independent assumption of characteristic conditions. For a given training data set, we first find out the combined input/output probability distribution based on the independent hypothesis of feature conditions. Then, based on this model, for the given input x, we use Bayes' theorem to find the output with the greatest subsequent probability y.

B. Unsupervised Learning

In this approach, the model works on itself by discovering or exploring the hidden patterns in the available information. Unlabeled data is associated only with the input data. This model figures out the relationship between data and divide it into clusters of the same patterns. It yield unpredictable results too and the main aim of this learning technique is to find the hidden patterns existing in the dataset. This learning is categorized into two parts: clustering and dimensionality reduction. K-means is the example of clustering and Principal Component Analysis (PCA) is example of dimensionality reduction.

1) Principal Component Analysis: This learning algorithm is quite popular into data compression field and is used for dimensionality reduction. It is a multivariate method and aims to extract important information from data in terms of principal components, which is nothing however a set of new orthogonal variables.

The data compression and dimensionality reduction is a multivariate method. It's objective extract crucial information from data. Also, it as a couple of new orthogonal variables knows as principal components. These principal components are ordered such that the first principal component is aligned in the direction of the highest-variance path of data, with reducing variation for additional components in order. This permits, the minimum variance components to be abandoned as they simply include least information content, causing dimensionality decrease. For WSN situations, this could lower the quantity of data becoming transmitted among sensor nodes by getting a tiny pair of uncorrelated linear blend of innovative readings. Further, it can solve the big data problem into small data by allowing selection of only significant principal components and discarding other lower order insignificant components from the model.

2) K-Means Clustering: This unsupervised learning algorithm classifies data into different clusters or classes and works in sequential steps involving, random selection of k nodes as initial centroids for different clusters, use of a distance function to instructions every node with the nearest centroid, iteratively re-compute the centroids using a predefined threshold value on present node memberships and quit the iterations if the convergence condition is met. The K-means clustering algorithm is popular in WSN sensor node clustering because of the simplicity and linear in its complexity.

C. Reinforcement Learning

It is an approach in which decision-making is carried out sequentially. The training of network is done intelligently in such a way that it automatically gives the best optimal action within a particular context or situation and thus increase efficiency and performance. The feedback-reward mechanism is mandatory for the agent and can be positive or negative. Positive reward increases the efficiency of the performance but on the other hand, negative reward degrades the performance of the network. This feedback reward is a reinforcement signal and it helps the agent to learn the network's behavior. The most promising example of this technique is the Q-Learning.

IV. CONCLUSION

Wireless sensor networks are different from traditional network in various aspects, thereby necessitating protocols and tools that address unique challenges and limitations. As a consequence, wireless sensor networks require innovative solutions for energy aware and real-time routing, security, scheduling, localization, node clustering, data aggregation, fault detection and data integrity. Machine learning provides a collection of techniques to enhance the ability of wireless sensor network to adapt to the dynamic behavior of its surrounding environment. This paper has provided a comprehensive review and analysis of machine learning techniques for wireless sensor networks (WSNs). Machine learning techniques have the potential to revolutionize the capabilities of wireless sensor networks, enabling intelligent data analysis, efficient resource utilization, and informed decision-making. By combining the strengths of machine learning algorithms with the unique challenges and requirements of WSNs, researchers and practitioners can unlock new possibilities for a wide range of applications and contribute to the advancement of this rapidly evolving field.

References

[1] S. Singh, K. Kumar, and S. Gupta, "Machine Learning based Indoor Localization Techniques for Wireless Sensor Networks", Proceedings -IEEE 2020 2nd International Conference on Advances in Computing, Communication Control and Networking, ICACCCN 2020, pp. 373–380, 2020, doi: 10.1109/ICACCCN51052.2020.9362802.
[2] T. O. Ayodele, "Introduction to machine learning," in New Advances in Machine Learning. InTech, 2010.
[3] Vikash Kumar, Anshu Jain and P N Barwal, "Wireless Sensor Networks: Security Issues, Challenges and Solutions", International Journal of Information & Computation Technology. ISSN 0974-2239 Volume 4, Number 8 (2014), pp. 859-868.
[4] Dr. Dattatray G. Takale, Prof. Vajid Khan, "Machine learning techniques for Routing in Wireless sensor network", 2023 IJRAR January 2023, Volume 10, Issue 1.
[5] Alsheikh, M.A.; Lin, S.; Niyato, D.; Tan, H.P., "Machine Learning in Wireless Sensor Networks: Algorithms, Strategies, andApplications", IEEE Commun. Surv. Tutor. 2014, 16, 1996–2018.
[6] M. Di and E. M. Joo, "A survey of machine learning in wireless sensor netoworks from networking and application perspectives," in 6th International Conference on Information, Communications Signal Processing, 2007, pp. 1–5.
[7] A. Forster, "Machine learning techniques applied to wireless ad-hoc networks: Guide and survey," in 3rd International Conference on Intelligent Sensors, Sensor Networks and Information. IEEE, 2007, pp. 365–370.
[8] A. Förster and M. Amy L, Machine learning across the WSN layers. InTech, 2011.
[9] Y. Zhang, N. Meratnia, and P. Havinga, "Outlier detection techniques for wireless sensor networks: A survey," IEEE Communications Surveys & Tutorials, vol. 12, no. 2, pp. 159–170, 2010.
[10] V. J. Hodge and J. Austin, "A survey of outlier detection methodologies," Artificial Intelligence Review, vol. 22, no. 2, pp. 85–126, 2004.
[11] R. Kulkarni, A. Förster, and G. Venayagamoorthy, "Computational intelligence in wireless sensor networks: A survey," IEEE Communications Surveys & Tutorials, vol. 13, no. 1, pp. 68–96, 2011.
[12] S. Das, A. Abraham, and B. K. Panigrahi, Computational intelligence: Foundations, perspectives, and recent trends. John Wiley & Sons,Inc., 2010, pp. 1–37.
[13] Kaur, T., Kumar, D., 2019. QoS mechanisms for MAC protocols in wireless sensor networks: a survey. IET Commun. 13 (14), 2045–2062.
[14] Aoudia, F.A., Gautier, M., Berder, O., 2018. RLMan: an energy manager based on reinforcement learning for energy harvesting wireless sensor networks. IEEE Transactions on Green Communications and Networking 2 (2), 408–417.
[15] Y. S. Abu-Mostafa, M. Magdon-Ismail, and H.-T. Lin, learning from data. AMLBook, 2012.
[16] K. Beyer, J. Goldstein, R. Ramakrishnan, and U. Shaft, "When is "nearest neighbor" meaningful?" in Database Theory. Springer, 1999, pp. 217–235.
[17] T. O. Ayodele, "Types of machine learning algorithms," in New Advances in Machine Learning. InTech, 2010.
[18] S. R. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology," IEEE Transactions on Systems, Man and Cybernetics, vol. 21, no. 3, pp. 660–674, 1991.
[19] Y. Bengio, "Learning deep architectures for AI," Foundations and Trends in Machine Learning, vol. 2, no. 1, pp. 1–127, 2009.