Integration of Machine Learning Techniques in Virtual Wireless Sensor Network for insect monitoring

Mohammad Equebal Hussain¹ and Rashid Hussain²

¹Department of Computer Science, Suresh Gyan Vihar University, Jaipur
²mdequebal.60508@mygyanvihar.com and rashid.hussain@mygyanvihar.com

Abstract. Agriculture industries are comparatively slow in adopting emerging technologies than any other industries despite lot of exciting research. The use of Wireless sensor network (WSN) is very important role in agriculture for more productive and sustainable growth. The structure of WSN is tightly application dependent. Every WSN have sensors, processing unit, low frequency radio wave transmitter and power supply using battery. With increasing number of interconnected WSN devices, there is substantial increase in data generation. It contains both control messages as well as application dependent data, collected by the sensors. The collected data are frequently sent to the nearest centralized controller for further processing and decision making. For continuous functioning of WSN, uninterrupted power supply is needed. Many researches are carried out to overcome these challenges. In this manuscript we are proposing a simple and effective machine learning techniques combined with pause and play method to increase battery life of WSN. This could be achieved in three stages play-pause-play (PPP) model. First by gathering data for some time (play mode), Second by putting WSN to sleep (pause mode), in the backend, apply machine learning algorithm that helps to build model from training data to predict the future data. At the final stage, put WSN back to play mode. Compare the result with actual data and fine tune the model to reduce the error. Using this method WSN will get enough sleep time to increase overall life by simulating the normal behaviour of sensor node. The sleep time will be calculated dynamically.

Keywords: virtual wireless sensor network (vWSN), Machine Learning, agriculture, insect monitoring, naive bayes.

1. Introduction

Wireless sensor network (WSN) refers to group of dispersed and dedicated sensors, in order to monitor and record various environmental conditions like weather, temperature, pollution, soil, insect monitoring and so on. WSN performs its task using limited memory, low bandwidth, minimal power supply and computational capacity for sensing, processing and communicating. During the lifetime of WSN, several control messages were exchanged such as session establishment HELLO message, routing update and query or command related message between base station and sensor node. But the main purpose of WSN is to transfer sensing data. Virtual wireless sensor network (vWSN) [14] is an approach to minimize cost and increase throughput by separating data and control plane. Using virtual WSN, the data will be generated in backend without depending upon the sensor nodes using machine learning technique. Though vWSN in agriculture [11] for insect monitoring is recently introduced approach therefore not many references are available in this area. In this manuscript we are proposing
play-pause-play (PPP) model, in which virtual data will be generated using ML technique when WSN is in pause mode. This could be achieved automatically by using the data generated during play mode to train the model during pause mode, this is discussed in section 4.2. In this paper we will apply our model on agricultural data to predict and compare the results. Though there exists many algorithms but our experiment is based on Bayesian naive bayes method using sklearn library and GaussianNB method.

We have used online drawing tool [13] to draw all the diagrams in this paper.

In the subsequent section we will explore few related work. Section 3 will give an overview of commonly used machine learning techniques and algorithms. In section 4, we proposed PPP design model. Finally we will conclude the paper and future work in section 5.

1.1. Insect monitoring in agriculture

The purpose of early detection of insect attack on agriculture [12] is a preventive model rather than curative model. Based on the result, the action can be taken much prior to actual insect attack. This would help in reducing the use of pesticide and harmful chemicals to control health hazard and pollution. Continuous insect monitoring using WSN is already proposed in many earlier researches. The aim of this research is to combine the power of ML and WSN to save power, increase overall lifetime and correctly predict insect attack using virtual WSN. This would be helpful in reducing the cost of WSN deployment, maintenance as well as increase in profitability.

1.2. Problem statement, challenges and research gap

- Wireless sensor network (WSN) collects data too frequently from various sensor nodes and transmit it over the network, hence continuous bandwidth availability is one of the major challenge.
- Monitoring is a continuous process. WSN is completely dependent on uninterrupted power supply through battery which is too expensive to maintain for longer duration.
- WSN doesn’t apply any intelligence to predict the data rather dependent on the sensor node to produce real data. A faulty sensor can produce invalid data which couldn’t be validated.
- Taking the advantage of powerful and intelligent ML algorithm, when integrated with WSN, we could save on power and bandwidth requirement by predicting the data very close to the actual data. This will be useful to minimize the cost and increase throughput.

2. Related work

To achieve energy efficiency in WSN, several models have been proposed.

Bandur D [1], discusses about various factor contributing energy consumption which includes protocol level optimization using optimal number of sensor nodes sufficient to provide the service without degrading the quality as well as increasing the overall lifetime of sensor node. Ning and Cassandras presents an approach to save energy by optimizing the idle listening using dynamic sleep time and low power listening (LPL) technique [2]. Optimization based on energy efficient routing protocol is proposed in [3]. In [4], few energy efficient routing techniques were discussed. Yuan and Li [5] explore the machine learning technique for fault detection in WSN due to various wear and tear. Noshad and Javaid [6] propose various ML based algorithm to detect fault due to data loss in WSN. Rashid H [7] proposed a solution for insect monitoring using WSN and image processing through MATLAB technique based on the colour detection. D Praveen [7], [8] did the survey on various ML techniques, its advantages and drawbacks in WSN context. Virtualization model is proposed for WSN in [9]. Various machine learning techniques suitable for agriculture using WSN are discussed in [10]. Smart farming using IoT and WSN is discussed in [15] [17].

3. Machine learning techniques for insect monitoring

Insects can be detected and monitored automatically using modern sensor based technologies. Several research have already been done as well as in progress to improve the system. The main focus of this
The paper is to review the state of art ML technology combined with WSN and virtualization to build a framework for detection and monitoring of pest attack on agriculture. This could be achieved by predicting the data rather than depending entirely on real sensor data. The identification can be done using various methods like infrared, audio and image based sensors [16].

![Figure 1. Machine learning](image1)

![Figure 2. Commonly used machine learning algorithms](image2)

Any algorithm has some fundamental limitations and in this case the prediction accuracy depends on (i) data quality, (ii) dependent variable {output} and independent variable {input} as well as (iii) chosen model. Machine learning algorithms are majorly divided among various categories mentioned below [5] as well as shown in figure 2.

### 3.1. Regression

It comes under supervised machine learning useful to predict sensor data. It is further classified into linear and non linear regression. Linear model is based on one degree straight line equation, $Y = mX + c$ Where X is independent variable and Y is dependent variable. Linear regression is most effective and simple ML algorithm. Gaussian regression and support vector machine regression comes under non linear category [6]. Logistic Regression allows analyzing set of variables and predicting the outcome.

### 3.2. Decision Tree

It splits the data into various set. It follows a tree structure from root to leaf node. The best path is having minimum cost metrics (Mean Squared Error or MSE).

$$MSE = \frac{1}{n} \sum_{i} (\hat{y} - y_{i})^2, \text{where } \hat{y} \text{ is predicted value and } y_{i} \text{ is actual value}$$ (1)

### 3.3. Bayesian model

It is most simple and powerful probability based model requires relatively less number of training samples. It is based upon the relationship between probabilities of hypothesis before and after getting the evidence represented by $P(H)$ and $P(H|E)$ respectively. Where $P(H)$ is called prior probability and $P(H|E)$ is called posterior probabilities.
Using this method, likelihood of each class is calculated using posterior probabilities. The class having maximum likelihood is considered as the result. For research purpose we use GaussianNB model available in sklearn library to apply on the agricultural dataset.

- Load dataset
- Create Naïve Bayes model
- Make predictions
- Prepare classification report
- Prepare confusion matrix

4. PPP model using Gaussian Naive Bayes

Gaussian NB is most popular, fast and easy ML algorithm to predict class of dataset. Naive Bayes works only with predictors that are categorical. Any numeric data need to convert into categorical variables before Naive Bayes Classifier [18] [19] [20] can be applied. This could be achieved using below steps.

- Converting dataset into frequency tables.
- Calculate individual probability to generate Likelihood table.
- Calculate posterior probability using Bayes theorem.

Python sklearn library has inbuilt support to do all the above mentioned steps using its API. The advantage of this method is that it requires training data to do estimation using equation (1). The goal is to maximize posterior hypothesis for the training dataset equation (3) by extending equation (1) as below.

\[
P(H|E) = \frac{P(E|H)P(H)}{P(E)} \quad \text{where } H \text{ is hypothesis and } E \text{ is evidence}
\]

\[
\text{posterior probability} = \text{prior probability} \times \text{likelihood ration}
\]

\[
P \left( \frac{H}{E} \right) = \text{Maximize} \left\{ \Pi_{k=1}^{n} P \left( \frac{H_{k}}{E_{k}} \right) \right\}
\]

Gaussian NB fits well to predict the pest attack because the pattern follows a normal distribution. Based on the multi class Naïve bayes classification can be used to predict the attack due to insect from the tuple described by attributes affecting the growth of specific type of insect. The sample data set are mentioned in table 1

In the classification process, the derived model is based on training data which accurately predict if pest attack happened or not. The target tuples are <Frequency (Hertz), Signal to noise ratio (decibels-dB), CO2 level (ppm)>

Table 2 doesn’t contain the complete table contains 300 rows. But it is only for reference.

| S.No | Frequency (Hz) | Signal2Noise Ratio(S/N) dB | CO2 level (ppm) | Classification (Critical, High, Medium, Low) |
|------|----------------|------------------------------|-----------------|---------------------------------------------|
| 1    | 700            | 0.7                          | 40000           | Critical                                    |
| 20   | 1100           | 1.1                          | 380             | Low                                         |
| 300  | 400            | 3.2                          | 3700            | Medium                                      |
4.1. Algorithm
- Import the dataset.
- Split into training set and test set (25% training set, 75% test set, factor = 0.25).
- Apply Gaussian naive bayes classifier to fit to the training dataset.
- Predict the result.
- Check the accuracy using confusion matrix.

Python sklearn library provides a facility to visualize the training and test dataset.

4.2. Proposed design

In the below figure 3, an intelligent play-pause-play model for vWSN is proposed with the use of machine learning. The WSN sensors will switch between play (on) and pause (off) mode at dynamically calculated interval. The data collected during play mode will be distributed among test and training set. When sensor is in paused state, ML algorithm will run to predict the data which can be compared with data saved in database. Finally the performance will be evaluated in figure 4 and figure 5.

![Play-Pause-Play (PPP) vWSN model](image)

**Figure 3.** Play-Pause-Play (PPP) vWSN model

4.3. Results

![Classification report](image)

**Figure 4.** Classification report

![Confusion Metrics](image)

**Figure 5.** Confusion Metrics

4.4. Metrics

Confusion metrics is a performance measure for classification problem in machine learning shown in figure 6 and figure 7.
5. Conclusion and future work

Like other industries, agriculture is also going through digital transformation. Wireless sensor network generates exponentially huge data. In this paper, we presented machine learning based PPP model for virtual wireless sensor network. In the proposed system, when sensor nodes are in pause mode, the data can be generated automatically as well as the system will keep improving its performance to correctly predict future data. In the future, we planned to extend the model such that it can be suitable for any type of WSN. We will also plan for field experiment to test the proposed design on various other ML algorithms.

Acknowledgements

I would like to thank Dr. Rashid Hussain for his encouragement and support during the research.

References

[1] Bandur D, Branimir Jakšić, Miloš Bandur, Srdan Jović. 2019 Jan;156:500-507. DOI: 10.1016/j.compag.2018.12.016.
[2] X. Ning and C. G. Cassandras, "Optimal dynamic sleep time control in Wireless Sensor Networks," 2008 47th IEEE Conference on Decision and Control, 2008, pp. 2332-2337, doi: 10.1109/CDC.2008.4738768.
[3] N. A. Pantazis, S. A. Nikolidakis and D. D. Vergados, "Energy-Efficient Routing Protocols in Wireless Sensor Networks: A Survey," in IEEE Communications Surveys & Tutorials, vol. 15, no. 2, pp. 551-591, Second Quarter 2013, doi: 10.1109/SURV.2012.062612.00084.
[4] S. Elsan and B. Hamdaoui, "A Survey on Energy-Efficient Routing Techniques with QoS Assurances for Wireless Multimedia Sensor Networks," in IEEE Communications Surveys & Tutorials, vol. 14, no. 2, pp. 265-278, Second Quarter 2012, doi: 10.1109/SURV.2011.020211.00058.
[5] Y. Yuan, S. Li, X. Zhang and J. Sun, "A Comparative Analysis of SVM, Naive Bayes and GBDT for Data Faults Detection in WSNs," 2018 IEEE International Conference on Software Quality, Reliability and Security Companion (QRS-C), 2018, pp. 394-399, doi: 10.1109/QRS-C.2018.00075.
[6] Noshad, Z., Javed, N., Saba, T., Wadud, Z., Saleem, M., Alzahrani, M., & Sheta, O. (2019). Fault Detection in Wireless Sensor Networks through the Random Forest Classifier. Sensors (Basel, Switzerland), 19.
[7] Hussain, Rashid. (2017). WSN CONTROLLED INSECTS MONITORING: IDENTIFICATION OF ONION THRIPS. International Journal of Current Advanced Research. 6. 4. 10.24327/ijcar.2017.5260.0684.
[8] Tanuja Jha.; Rashid Hussain.IJCAR, 2014,6,5257-5260
[9] D. Praveen Kumar, Tarachand Amgoth, Chandra Sekhara Rao Annavarapu, https://doi.org/10.1016/j.infus.2018.09.013
[10] Hussain M E, Quayyum M, Hussain M R and Hussain R, 2019,  (IJCNIS), ISSN: 2073-607X
[11] Mekonnen, Y.; Namaduri, S.; Burton, L.; Sarwat, A.; Bhansali, 2020, 167, 037522
[12] Saleh, Y., Sani Yahya, M., Abdul Dalyop, I., & Hussain, R. (2018). 7(3.36), 121-126. doi:http://dx.doi.org/10.14419/ijet.v7i3.36.29091
[13] online DRAW.IO. http://draw.io/
[14] I. Khan, F. Belqasmi, R. Glitho, N. Crespi, M. Morrow and P. Polakos, in IEEE Communications Surveys & Tutorials, vol. 18, no. 1, pp. 553-576, Firstquarter 2016, doi: 10.1109/COMST.2015.2412971.
[15] Mekonnen, Y, Lamar Burton, Arif Sarwat, and Shekhar Bhansali, “Iot sensor network approach for smart farming: An application in food, energy and water system,” In 2018 IEEE Global Humanitarian Technology Conference (GHTC), IEEE, 2018
[16] Konstantinos Liakos, Patrizia Busato, Dimitrios Moshou, Simon Pearson, and Dionysis Bochtis, “Machine learning in agriculture: A review,” Sensors, 18(8), 2674 (2018).
[17] S. Arul Mozhi Varman, Arvind Ram Baskaran, S Aravindh, and E. Prabhu,”Deep learning and iot for smart agriculture using wsn,” In 2017 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), IEEE, 2017
[18] Zwartjes, G. J. (2017). Adaptive Naive Bayes classification for wireless sensor networks. Universiteit Twente. https://doi.org/10.3990/1.9789036542630
[19] Shailendra Mishra and Mayank Singh, “An improved Energy Efficient Communication Protocol (IEECP) for Wireless Sensor Networks,” International Journal of Engineering and Advanced Technology (IJEAT), vol. 8, no. 4c, April 2019.
[20] Barnawi, A.; Keshta, I.M. Energy management in wireless sensor networks based on naive bayes, MLP, and SVM classifications: A comparative study. J. Sens. 2016