Literature Study

Michael Matusowsky, 11093804

July 2, 2018

Introduction

Sensors are devices used in everyday objects such as mobile phones, touch-sensitive elevators and self-driving cars. There are an innumerable amounts of application where sensors are used to gather data about the environment and then used in other electronics to monitor or react to the conditions being sensed such as the lighting in a room being adjusted based on the required illumination level or underfloor heating being activated or deactivated based on the current sensed ambient temperature. A problem arises in wireless sensor networks (WSN) where, due to the nature of wireless communication, data can be lost or corrupted during transmission due to external factors such as solar radiation corrupting data from satellites or congestion in a network causing packet loss when transmitting data to other devices. Furthermore the cost of implementing or needing to replace many physical sensors in a network can become prohibitively expensive.
This project will look at designing and implementing a wireless sensor network that will make use of data imputation methods and machine learning to realise virtual sensors that can completely replace physical sensor nodes and give accurate substituted data in place of nodes with failed sensor modules. To be able to design the proposed virtual sensor system, it is necessary to understand the principles of machine learning, virtual sensors and wireless communication in sensor networks. This literature study will look at some of those principles in the following sections.

Imputation methods

Missing data imputation is an important task that must be carried out in cases where it is crucial to use all the data that is available and to not discard entries that may have incomplete data. [1] and [2] evaluates the performance of three statistical methods and three machine learning methods to predict the recurrence of cancer in breast cancer patients where it is shown that the machine learning methods outperform the statistical methods. The statistical methods are mean imputation, hot-decking and the expectation maximisation algorithm. The machine learning methods are the k-NN algorithm, artificial neural network and self-organising map. Only the machine learning methods will be discussed further in terms of application in temperature virtual sensors.

k-NN imputation

The k-NN method of machine learning is a "lazy learning" algorithm that selects a node in a dataset and compares it with k neighbours in a certain radius around the node before deciding which type of data it belongs to as shown in figure 1 below.

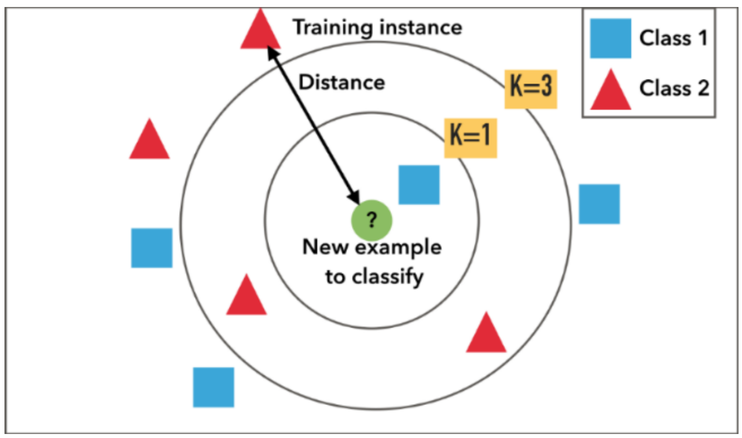

Figure 1: How the k-NN algorithm works.[3]

[4] implemented the k-NN algorithm to analyse and predict temperature and humidity from incoming data from sensors set up in the environment. The accuracy attained using this method of temperature and humidity prediction varies depending on the monthly dataset. This method has its shortfalls in that the predictions are strictly reliant on the available data such that outlier values might be misclassified by the algorithm due to lack of similar data near the outlying data node. However it is a computationally cheap and easy algorithm to implement.

Multi-Layered Perceptron Artifical Neural Network

The Multi-Layered Perceptron artificial neural network (MLP ANN), or vanilla neural network, is the basic form of the neural networks. They function by having one or more nodes that are fully connected in layers. The minimum number of layers consist of one input layer, one hidden layer and one output layer as shown in figure 2 below. MLP's are capable of finding relationships between linear and non-linear datasets making them useful as prediction models. Neural networks have been used to find the correlation between environmental factors and temperature for various reasons such as the implementations in [5] and [6]. [5] shows that by increasing the number of neurons and hidden layers,

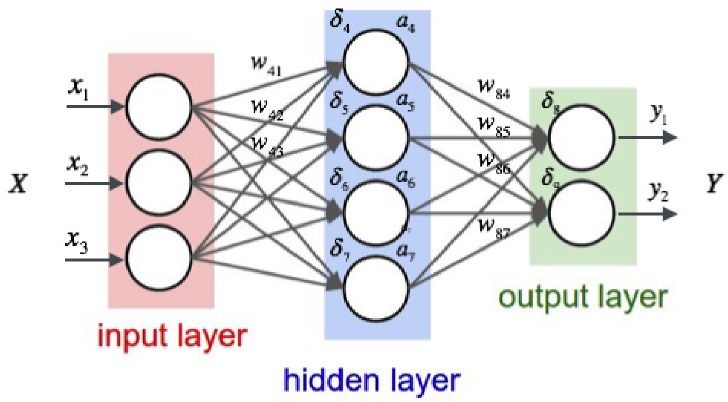

Figure 2: How a basic neural network is structured [7].

the performance of the atmospheric prediction increases. They train the neural network using the Levenberg-Marquardt algorithm [8][9] which is used to solve non-linear least squares problems. [6] makes use of the simple back-propagation algorithm as well as the Levenberg-Marquardt algorithm to determine the faster training method as well as to determine the optimal nodes and neurons for the neural network where it is found that the back-propagation algorithm performs better by requiring less epochs before settling at an acceptable error rate.

Self-Organizing Maps

The self-organizing map [11], shown in figure 3, is a type of artificial neural network developed by Teuvo Kohonen that makes use of an unsupervised learning algorithm to produce a discretised representation, typically in a two-dimensional map, of the input space of the training samples.
They differ from other neural networks in that they employ competitive learning [10] and using nearest neighbours to conserve topological properties. [12] makes use of this neural network to help visualise the complex relationships between climate data and temperature on a global scale. The map uses many different inputs such as maximum and minimum temperatures as well as height above sea level, relative humidity and barometric pressure. Once fully trained, the map is able to determine which weather stations were most likely to record a certain weather pattern to a very high degree of accuracy. The advantage of the self-organizing map is that the training algorithm is much simpler than the implementation required for the multi-layered perceptron and does not take as long to train. However it is limited in that it is an iterative process when deciding how many nodes in the neural net should be in place and due to the random initial state of the map, it is hard to find an optimal solution.

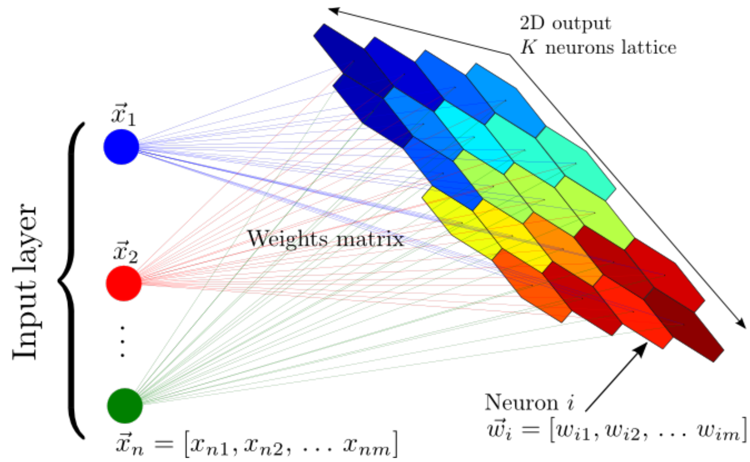

Figure 3: How self-organizing map is structured [12].

Conclusion

There are many different methods to finding the relationship between temperature and the surrounding atmosphere. The more complex methods, namely the artificial neural network and self-organzing map, tend to be more accurate than the simple k-nn algorithm but the difference is negligible. Due to the ability of the multi-layered perceptrons ability to find relationships in linear and non-linear data in a large amount of cases it is the desired machine learning method when developing predictive models as the amount of neurons and layers added is not a complex task as is the case in the self-organizing map where every output node is fully connected to every input node.

References

[1] J. Jerez, I. Molina, et al,“Missing data imputation using statistical and machine learning methods in a real breastcancer problem,” *Artificial Intelligence in Medicine*, vol. 50, no. 2, pp. 105–115, 2010.

[2] O. Troyanskaya, M. Cantor, et al, ”Missing value estimation methods for DNA microarrays,” *Bioinformatics*, vol. 17, no. 6, pp. 520-525, 2001.

[3] A. Bronshtein, ”A Quick Introduction to K-Nearest Neighbors Algorithm”, 2017. Available: https://medium.com/@adi.bronshtein/a-quick-introduction-to-k-nearest-neighbors-algorithm-62214cea29c7. [Accessed 14- June- 2018].

[4] S. S. Badhiye, N. U. Sambhe, et al, ”KNN Technique for Analysis and Prediction of Temperature and Humidity Data,” *International Journal of Computer Science*, vol. 61, no. 14, pp. 9-13, 2013

[5] K. Abhishek, M. P. Singh, et al, ”Weather forecasting model using Artificial Neural Network”, *Procedia Technology*, vol 4, pp. 311-318, 2012.

[6] S. Papantoniou, D. Kolokotsa, ”Prediction of outdoor air temperature using neural networks: Application in 4 European cities”, *Energy and Buildings*, vol. 114, pp. 72-79, 2012.

[7] V. Valkov, ”Creating a Neural Network from Scratch”, 2017. Available: https://medium.com/@curiousily/tensorflow-for-hackers-part-iv-neural-network-from-scratch-1a4f504dfa8 [Accessed 28- June- 2018].

[8] K. Levenberg, ”A Method for the Solution of Certain Non-Linear Problems in Least Squares”, *Quarterly of Applied Mathematics*, vol.2, pp. 164-168, 1944.

[9] D. Marquardt, ”An Algorithm for Least-Squares Estimation of Nonlinear Parameters”, *SIAM Journal on Applied Mathematics*, vol. 11, no. 2. pp. 431-441, 1963.

[10] D. Rumelhart, D. Zipser, et al, ”A Competitive Learning Procedure”, *Parallel Distributed Processing*, vol. 1, pp. 151-193.

[11] T. Kohonen, E. Oja, et al, ”Engineering Applications of the Self-Organizing Map”, *Proceedings of the IEEE*, vol. 84. no. 10, 1996.

[12] L. Haihan, ”Analyzing Climate Patterns with Self-Organizing Maps”, Available: https://towardsdatascience.com/analyzing-climate-patterns-with-self-organizing-maps-soms-8d4ef322705b [Accessed 28- June- 2018].