Visualization and Statistical Analysis of Multi Dimensional Data of Wireless Sensor Networks Using Self Organising Maps

Thendral Puyalnithi[#1], V Madhu Viswanatham[*2]
School of Computer Science and Engineering,
VIT University, Vellore-632 014, Tamilnadu, India.
[1]thendral.p@vit.ac.in,thendral_psg@yahoo.com
[2]vmadhuviswanatham@vit.ac.in

Abstract --- **A Wireless Sensory Network(WSN) has put much importance to the data that it gets from the Sensor nodes, since the data from sensors are so valuable since sensor has limited life time and moreover the sensors are mostly not reachable physically once they are deployed, so an efficient analysis of the valuable data needs to be place once the data is obtained , so that decision making is can be effective. The Self Organizing Maps(SOM) technique of Artificial Neural Network, which are proven methodologies for clustering the data has more advantage and it has been proposed in this paper for the sink data analysis and the SOM has been modified, by storing the sensor location information in its Best Matching Unit, so that SOM can be used for Classical Statistical Analysis too.**

Key Words -- Wireless Sensor Network, Artificial Neural Network, Unsupervised Learning, Self Organising Map, Clustering

I. INTRODUCTION

Wireless sensor networks are vital components in Military applications, Environmental applications, Health applications, Home applications etc. For example Wireless Sensor Network (WSN) nodes deployed over certain terrain can be used to detect temperature in various locations in that area. The Network then sends the data to the *sink* node. Every *sink* nodes data are analyzed. The sensor networks data are volatile in nature, so for every time interval, the data will be got from the sensor networks for analysis. Every node in sensor network not collects just single dimensional data, ie one dimensional data, it collects multi dimensional data. The multidimensional data creates much more difficulties in analyzing the data in the *sink* node side, so application of unsupervised machine learning technique of Self Organising Maps is proposed for the analysis of *sink* node data.

II. WIRELESS SENSOR NETWORKS (WSN)

A. WSN Network

Advancements is microelectronics, digital electronics, mechanical systems, wireless systems along with various soft computing methods made the sensor networking technologies easily deployable and usable in various terrains. Wireless Sensor network comprises of small sensor nodes. The nodes are deployed in large number in a certain terrain, for example in some environment it can even be deployed to the extent of millions. The sensor network nodes are alive only till the power system in it depletes, so sensor network nodes must be very much power aware. Sensor nodes can sense various parameters in an area, like temperature, humidity, vehicular movement, pressure, soil makeup, noise levels, presence or absence of certain kind of objects, mechanical stress levels on attached objects, speed and size of an object. The Job objective of each sensor network's node is to do Sensed Data Analysis, Data Transmission along with Power Management. It has to process the data, it has sensed. It has to send the processed data to other nodes based on certain wireless sensor network protocol. Sensor network can be deployed in the terrains which are hazardous to reach by human, for example Sensor network can be deployed inside volcanoes and the activities inside the volcanoes can be monitored, so that the warning systems can be fed with those data.

B. WSN Challenges

There is no fixed topology or infrastructure for WSN, i.e. Nodes do not have any a priori knowledge about the rest of the network. Hence A WSN must be able to self-organize. Nodes cooperate for this self-organization and also along the whole life of the network to achieve the requested service.

All the WSN nodes will not be alive all the time. Every node will become 'on' and 'off' in a certain time gap, to save power, but some sensor nodes has solar panel to generate power, the decision of integrating power generation unit along with the sensor node is application oriented. The topology of the deployed network will not be a fixed as in normal sensor networks, since sensor nodes' status will be changing from alive to dead and vice versa frequently, moreover sensor nodes in certain application can change its location also, for example sensor nodes deployed in sea will be changing its position as per the current of water in that area.

Every node in WSN has to send its data to a node called '*Sink* Node'. Every node follows multi hop approach to send its data to the *sink* node since they cannot communicate directly with their radio. Every node sends its data to the neighbouring nodes in many ways, for example a node can follow flooding approach. It can flood the data it has, to its neighbours and in-turn the neighbours will also flood, by doing so the sensed data of every node in the WSN will reach the *sink* node.

C. Sink Node in Sensor Networks

Sensor networks group together and they form a topology, in that topology the sensed data will be aggregated and sent to the *sink* node. The *sink* node is the one which is connected to internet or satellite and the *sink* nodes will have good power back up.

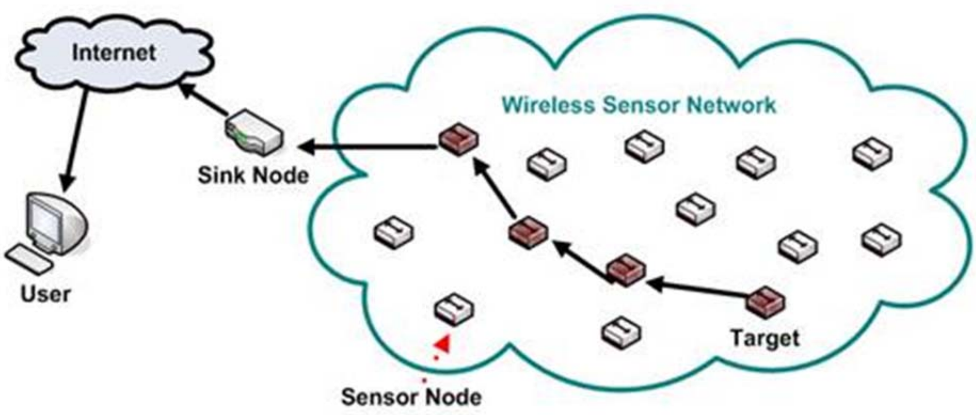

Fig 1. General Wireless Sensor Network Model

Sometimes the *sink* node will also be not fixed, meaning when the topology changes the *sink* node can also change. In other works we can say that the node which is connected to internet or which gets all the data from all the other nodes in the sensor network is said to be *Sink* Node. The Fig 1[Ref. 7]. Shows a general wireless sensor network model the sink node gets the stream data from wireless sensor network.

D. Issues in Analysis of Sink Node Data

Sensor Networks' sensed data will be stored in its *Sink* nodes as in Fig 2[Ref.3].

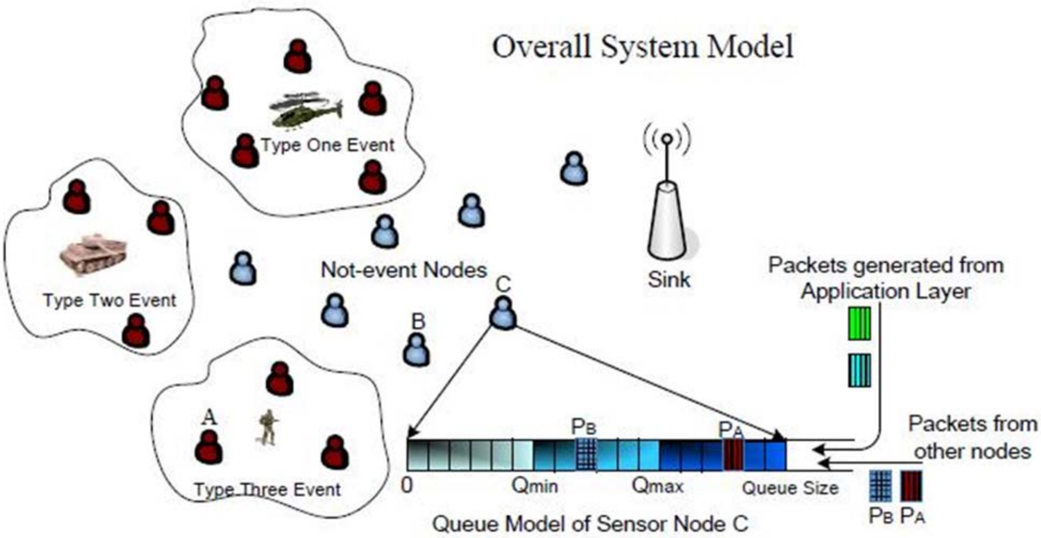

Fig 2. Wireless Sensor Network Data Streaming

As mentioned earlier the data obtained from the sensor nodes are multi dimensional. For example sensor node deployed in a terrain to do environmental studies will record the parameters such as temperature, humidity and time whenever it senses. The sensor goes on recording this three dimensional data and sends the collected data to the *sink* node. Thus a *sink* node will be having collection of this three dimensional data.

In a scenario where a sensor network is deployed in forest and it records temperate, humidity and time every time it senses. The *sink* node will obtain this data from sensor nodes. If an attribute based search is imposed on the data, for example we want to know how much percentage area in the forest has humidity more than certain value and temperature below certain value during certain month, and then linear analysis through the data will fetch the answer. But if we need to cluster the multidimensional data, then Linear Analysis will not be efficient.

III. PROPOSED METHOD

A. Applying Self Organizing Maps (SOM) for Sink Node Data Analysis

In this paper we have proposed to use Kohonen Self Organising Maps (SOM) for clustering the *sink* node data. The reasons for going for SOM are

a) It is an unsupervised Neural Network learning methodology, since we want to predict the things based on the information got from sensor networks, we need this characteristics.

b) SOM coverts high dimensional data into low dimensional data, so that we can visualize easily.

c) In case if you need coverage analysis on certain data then this is the proven analysis method for clustering.

d) The classical statistical analysis will do Attribute based search only, but SOM will do the clustering analysis of the entire data set.SOMs clusters allows user to visualize the homogeneous data groups.

SOM algorithm is applied for cluster analysis of the *sink* data. Here we feed the SOM with the high dimensional data obtained from the *sink* nodes. The some will organize the data into clusters. The boundary analysis of the clusters gives the cluster size and then the volume distribution of data can be obtained.

B. Imposing Classical Statistical Analysis from SOM

As of now we have been planning to categorize the sensor nodes' data into clusters, but if we want to do statistical analysis, then we have store location dimension data in the BMU unit of SOM. By doing this we are associating the location of sensors with the BMU that paves the way for doing classical statistical analysis.

C. SOM Algorithm

The SOM algorithm for clustering is given below

Step 1 : Each node's weights are initialized.

Step 2 : A vector is chosen at random from the set of training data and presented to the network.

Step 3 : Every node in the network is examined to calculate which ones' weights are most like the input vector. The winning node is commonly known as the Best Matching Unit (BMU).(Equation 1).

Step 4 : The radius of the neighbourhood of the BMU is calculated. This value starts large. Typically it is set to be the radius of the network, diminishing each time-step. (Equation 2a, 2b).

Step 5 : Any nodes found within the radius of the BMU, calculated in 4), are adjusted to make them more like the input vector (Equation 3a, 3b). The closer a node is to the BMU, the more its' weights are altered (Equation 3c).

Step 6 : Repeat 2) for N iterations.

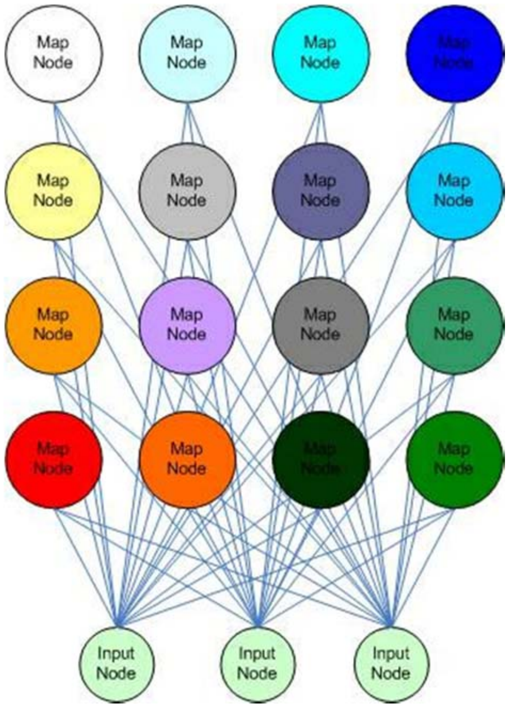

Fig 3. Neural Network node initial connections, each node has random colour value(Ref.[8])

Equation 1 – Calculate BMU.

$$\textbf{\textit{DistFromInput}}^2 = \left(\sum_{i=0}^{i=n} |I_i - W_i\right)^2$$

I – Current Input Vector

W – Node's Weight vector

N – Number of weights

Equation 2a – Radius of the neighbourhood.

$$\sigma(t) := \sigma_0 e^{-t/\lambda}$$

t – Current iteration

λ – Time constant (equation 2b)

σ_0 – Radius of the map

Equation 2b – Time constant.

$$\lambda = numIterations / mapRadius$$

Equation 3a – New Weight of node.

$$W(t+1) = W(t) + \theta(t)L(t)\big(I(t) - W(t)\big)$$

Equation 3b – Learning Rate.

$$L(t) = L_0 e^{\left(\frac{-t}{\lambda}\right)}$$

Equation 3c – Distance from BMU

$$\theta(t) = e^{\left(\frac{-distFromBMU^2}{2\sigma^2(t)}\right)}$$

E. Experiment and Results

We have simulated the experiment by generating the temperature data randomly to illustrate the algorithm. Let us consider that sensor network deployed in a dense forest floor to analyze the change climatic conditions at the ground level. Temperature, humidity, light intensity, time of sensing and location are sent as data to the *sink* Node. The vector has 5 scalar data as mentioned above. In the below figure every coloured box in a row indicates a high dimensional data obtained from the sensor node. This data is fed to the SOM and we got the cluster analysis of data as in figure

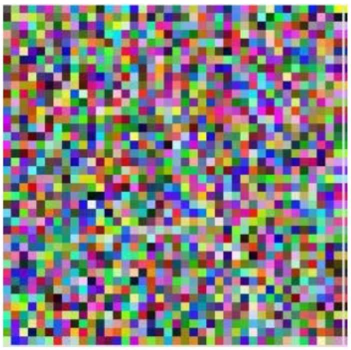

Fig 4. Random Initial Node values

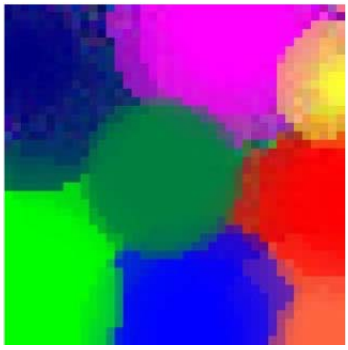

Fig 5. Clustered Values after algorithm execution

Thus we come to know the extent of diversification of temperature, humidity, light intensity in the entire forest region, but the above cluster analysis method will not give the location sensor node. To solve this issue we have to store the location of sensor node data obtained as part of high dimensional data in the Best Matching Unit (BMU) in the SOM, then we will be able to map the location of sensors in the cluster obtained. By doing this we will be getting the cluster, and the location of sensors which are part of that clusters. Thus classical statistical analysis is also added to the visual analysis.

Therefore from the SOM analysis, Prediction of environmental changes can be done. In an application of a kind where the nearby sensors give data of less deviation, then those sensors will be as part of same cluster in the SOM. So while we are getting the location information of sensors in the cluster, if we get a sensor which is not physically located others in the cluster, then we will come to know the existence of event which is not similar to others in its locality. Based on SOM cluster analysis we can also come to know the extent of further deployment of sensor nodes to locations in the forest floor.

IV. CONCLUSION

Thus SOM can be effectively used for cluster and statistical analysis of wireless network stream data. But the major issue in WSN is changing of sink nodes time to time, so SOM maps has to be generated repeatedly or the sink nodes' data has to be backed up and the new sink nodes' data has to be merged with backed up one to come out with the new Self Organising map.

REFERENCES

[1] Jennifer Yick, Biswanth Mukherjee and Dipak Ghosal, "Wireless Network Survey" doi:10.1016/journal.comnet.2008 .01.0002
[2] I.F. Akyildiz, W.Su, Y.Sankarasubramaniam, E.andCayirci " Wireless Sensor Network : a survey" , Elsevier Science. B.V, 2002
[3] XiaoligQiu, HaipingLiu,DeshiLi,JenniferYick, Dipak Ghosal and Biswanth Mukherjee, "Efficient Aggregartion of Multiple classes of Information in Wireless Sensor Networks", Sensors,2009,9,8083-8108;doi:10.3390/s9100803
[4] J.Eggermont, "Rule Extraction and Learning in BP-SOM Architecture" , Leiden University, Internal Report IR-98-16
[5] ShyamGuthikonda, "Kohonen Self Organizing Maps", Wittenburg University, Internal Report
[6] Cerpa, D. Estrin, "ASCENT: adaptive self-configuring sensor networks topologies", UCLA Computer Science Department Technical Report UCLA/CSDTR-01-0009, May 2001.
[7] http://wsnbydefault30.blogspot.in/2013/10/wireless-sensor-network.html
[8] http://www.generation5.org/content/2004/aisompic.asp

AUTHOR PROFILE

Mr.Thendral Puyalnithi has done his Bachelor of Engineering in PSG College of Technology, Coimbatore, Tamilnadu, India and Master of Engineering in BITS Pilani, Rajasthan, India. Currently he is working as Assistant Professor Senior in School of Computer Science and Engineering, VIT University, Vellore, Tamilnadu, India and pursuing PhD. His research area is in Data mining and Data analytics.

Dr.V MadhuViswanatham has done his doctorate in the field of network security and data security. Currently he is working as Associate Professor in School of Computer Science and Engineering, VIT University, Vellore, Tamilnadu, India. His research area is in Network data security, Data mining and Data analytics.