International Journal of Energy, Information and Communications
Vol.6, Issue 2 (2015), pp.23-32
http://dx.doi.org/10.14257/ijeic.2015.6.2.03

Energy Conservation in Wireless Sensor Networks by Exploiting Inter-Node Data Similarity Metrics

Sunil Dhimal[1] and Kalpana Sharma[2]
Sikkim Manipal Insitute of Technology,
Sikkim Manipal University
[1]*sunildhimal@gmail.com*, [2]*kalpanaiitkgp@yahoo.com*

Abstract

Wireless Sensor Networks (WSN) is the collection of large number of low powered sensor nodes deployed for sensing & monitoring real world physical phenomenon like temperature, humidity, soil moisture, pressure, etc. Due to low powered battery and unattended deployment of sensor nodes, conserving sensor node energy requirement is one of the prime challenges in WSN. Satisfactory coverage of sensing area requires sensor nodes to be deployed densely which causes the real time observation to be associated spatially or temporally. Due to such dense deployment, there may be a situation where in more than one sensor node captures and communicates the same physical phenomenon which hints that data aware energy conservation technique may be applied in order to reduce the overall power requirement of the network.

In this paper, a model is proposed to conserve overall energy by exploiting inter-node data association. In order to achieve this, real time wireless sensor network scenario is deployed and spatially and temporally associated sensor node data is acquired for further analysis. The inter-node data similarity is then measured by calculating similarity metrics like -Euclidean Distance, Cosine similarity and Pearson correlation coefficient. Finally, energy conservation technique is applied if the similarity metrics suggests high correlation between sensor nodes data.

Keywords: *WSN, Similarity Metrics, Euclidean Distance, Cosine similarity, Pearson correlation coefficient*

1. Introduction

WSN has an immense scope in designing an application for remote sensing, monitoring and analyzing parameters of interest like temperature, humidity, pressure, seismic wave, soil moisture etc. which helps in better decision making. It is the known fact that energy conservation is one of the major challenges in wireless sensor nodes owing to their limited battery life and unattended operations [1]. Therefore, major concern in WSN is to reduce the overall energy consumption. Typical WSN applications require dense sensor deployment in order to achieve satisfactory coverage as mentioned in [10]. As a result, multiple sensors record information about a single event in the sensor field. Due to high density in the network topology, spatially proximal sensors observations are highly correlated with the degree of correlation increasing with decreasing inter node separation [2, 10].

Although various media access control(MAC) and routing schemes has been proposed for energy conservation, no promising techniques has been proposed in the direction of energy efficient data acquisition[3]. There are various data-driven techniques that are implemented to conserve energy – like in-network processing and data compression techniques. In-network processing mainly focuses on performing data aggregation at intermediate nodes in order to reduce communication. Data compression technique is used in order to reduce the size of data to be sent between sensor nodes [3].

ISSN: 2093-9655 IJEIC
Copyright ⓒ 2015 SERSC

This paper proposes data-aware energy conservation scheme which primarily reduces energy consumption by preventing redundant communication which is identified at base station by performing various statistical and mathematical analysis on datasets generated by sensor nodes. The inter-node data similarity is measured using similarity metrics like Euclidean Distance, Cosine Similarity and Pearson correlation coefficient. A data-aware algorithm is designed; the working of algorithm is verified by performing simulation on 25000 temperature, humidity and soil moisture datasets generated by two different wireless sensor network applications. Once high correlation between inter-node datasets is found, the redundant sensor node is switched to sleep mode in order to reduce redundant sensing and communication. By switching from active to sleep mode energy is conserved [12].

The remainder of this paper is organized as follows. Section 2 of this paper briefly highlights the related work in this field. Section 3 presents the details about the similarity metrics like Euclidean Distance, Cosine Similarity and Pearson Coefficient. Section 4 is dedicated towards algorithm design for energy conservation based on similarity metrics. Section 5 discusses about the implementation and datasets that is required for verification of the purposed algorithm. Section 6 talks about analysis and result. Section 7 discusses the results and Section 8 concludes the paper.

2. Related Work

The data-driven energy conservation techniques as appeared in literature are discussed below:

In accordance to paper [3], data-driven approaches to conserve energy are divided into **data reduction schemes** and **energy-efficient data acquisition schemes**. Data reduction approaches address the case of unneeded samples, while data acquisition approaches are mainly aimed at reducing overall energy spent for communication as well. It also highlights that in-network processing, data compression and data forecasting techniques are most widely used data-driven techniques for overall energy conservation.

The spatial and temporal correlation along with collaborative nature of WSN can be exploited in order to propose efficient communication protocol well suited for WSN paradigm. A theoretical framework for the same in proposed in [10]. There also has been an effort made to study spatial and temporal association and analyze them at MAC and Physical layer.

The literature [2] proposes a methodology that uses computational geometry (exact and greedy approach) to identify spatially and temporally related sensor nodes from the set of all sensor nodes. The real time verification is conducted taking help of metrics like Jaccard's coefficient, Cosine similarity and Pearson correlation coefficient.

As mentioned in paper [7], adaptive power mode switching makes optimal power consumption. The energy component of radio like switching, transmission, reception, listening and sleeping are major contributing factors.

3. Similarity Metrics

In this section the theoretical methods to calculate inter-node data association is discussed as mentioned in paper [3]. Let $x[i]$ be the datasets from sensor node X and $y[i]$ be the datasets from sensor node Y. Let n be total number of samples used for analysis. The commonalities between data source due to spatial and temporal relation can be calculated using following similarity metrics.

 Copyright ⓒ 2015 SERSC

3.1. Euclidean Score

Euclidean score is a method of calculating a score of how similar two things are. We get a value between 0 and 1, 1 meaning they are identical 0 meaning they don't have anything in common. Euclidean score is calculated as given in equation 1

$$Euclidean\ Score = \frac{1}{(1 + Euclidean\ Distance)} \qquad (1)$$

Euclidean distance is the distance between two points in Euclidean space. Euclidean space was originally devised by the Greek mathematician Euclid around 300 B.C.E. to study the relationships between angles and distances. The Euclidean Distance is calculated using following formula in equation 2

$$Euclidean\ Distance = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \qquad (2)$$

3.2. Cosine Similarity

The cosine similarity between two nodes is given by equation 3

$$Cosine\ Similarity(\cos\theta) = \frac{\sum_{i=1}^{n} x_i \times y_i}{\sqrt{\sum_{i=1}^{n} {x_i}^2} \times \sqrt{\sum_{i=1}^{n} {y_i}^2}} \qquad (3)$$

The cosine similarity is in the range of [−1,1], the similarity value of −1 means that data are exactly opposite, 0 meaning independent,1 meaning exactly the same, with in-between values indicating intermediate similarities or dissimilarities. Cosine similarity gives a measure of common variations observed between the samples of the nodes by evaluating the angles between the two data samples considered.

3.3. Pearson Correlation Coefficient

Pearson Correlation Coefficient*(r)* is known as Pearson Product-Moment Correlation Coefficient (PPMC). The Pearson product-moment correlation coefficient is a measure of the linear correlation between two variables X and Y, giving a value between +1 and −1 inclusive, where 1 is total positive correlation, 0 is no correlation, and −1 is total negative correlation. As the value of *r tends to 0*, the greater is the variation. The formula for calculating PPMC is given in equation 4

$$PPMC(r) = \frac{n \sum xy - \sum x \sum y}{[\sqrt{n \sum x^2 - (\sum x)^2}]\,[\sqrt{n \sum y^2 - (\sum y)^2}]} \qquad (4)$$

4. The Algorithm

The algorithm that is designed to estimate inter-node data association using similarity metrics is proposed in this section. It is assumed that the datasets generated by sensor nodes are related to some degree due to dense sensor deployment and temporal relationship between inter-node sensor data. The algorithm discussed below is application specific and certain mission-critical application may not benefit from this approach.

The algorithm runs on the base station of wireless sensor network with the datasets that is discussed in the Section 4. The algorithm finally calculates the number of samples for which specified degree of data similarity is found thereby suggesting implementation of energy conservation scheme. For the sensor nodes available within the same cluster, it collects datasets (s_size) for some physical phenomenon and calculates inter-node data similarity. Once the inter-node data similarity metric is sufficiently found out to be

Copyright ⓒ 2015 SERSC

similar, the redundant sensor node is switched to low power consumption mode for some duration so that energy is conserved.

The algorithm is given below:

//Initialize variables

s_size (the number of samples for each iteration of analysis)
skipped_samples(total redundant samples, initially set to 0)
threshold_euclidean_score(threshold for Euclidean score)
threshold_pearson_coefficient (threshold for PPMC)
threshold_cosine_similarity (threshold for Cosine similarity)

While (Sensor Node System is switched ON)
{
For each sensor nodes (X, Y) in same cluster
{
While (Data samples available)
{

1. *Collect s_size datasets x[i] and y[i] from sensor nodes X & Y respectively.*
2. *Calculate similarity metrics [Pearson Coefficient{r}, Cosine Similarity {Cos (theta)}, Euclidean Distance {d(x,y}] for finding inter-node data association.*
3. *Does similarity metrics indicate high correlation?*
4. *If similarity metrics indicate high correlation between X & Y (i.e. r > threshold_pearson_coefficient or Cos (theta) > threshold_cosine_similarity or Euclidean score > threshold_euclidean_score) Then*
    a. *Switch power mode of X or Y nodes (say X) to sleep for next s_size/2 samples.*
    b. *skipped_samples = skipped_samples + s_size/2*

}
}
Display total samples skipped (skipped_samples) for (X, Y) due to redundancy
Calculate energy conservation percentage from skipped_samples
}

There are certain parameters that have to be configured and modified as per application demands. The list of configurable parameters is given below in Table 1 along with its value for an experiment setup conducted for a sensor node scenario implemented as discussed in Section 4.

Table 1. List of Configurable Parameters Along with its Values

S. No	Parameter	Value
1	s_size	20
2	threshold_euclidean_score	0.75
3	threshold_pearson_coefficient	0.99
4	threshold_cosine_similarity	0.99

5. The Program and Datasets

The Similarity metrics formula discussed in Section 2 and Algorithm discussed in Section 3 is used to design a program in C programming language in order to verify the concept. The first set of data for the program is taken from the Labelled Wireless Sensor Network Data Repository (LWSDR) of The University of North California

 Copyright ⓒ 2015 SERSC

(GREENBORO) [3]. The datasets from LWSDR is a Multi-hop labelled temperature and humidity data generated by 4 sensor nodes. The hardware platform that we used for data collection was the TinyOS-based motes, more specifically, Crossbow TelosB motes

As shown in Figure 1, there is a base station node with ID 0 as the root of the tree and two router nodes with the IDs of, respectively, 10 and 20. Also, there are two indoor sensor nodes with respective ID numbers of 1 and 2 and two outdoor sensor nodes (3 and 4). Each sensing node has a reduced transmission range and cannot reach the base station node except through the router node. The sampling interval is 5 seconds and each packet has ten readings each of temperature and humidity. The time period for collecting the readings is 6 hours, during which anomalies were introduced to one sensor node in each scenario (nodes 1 and 3) by using a water kettle which altered the temperature and humidity appropriately.

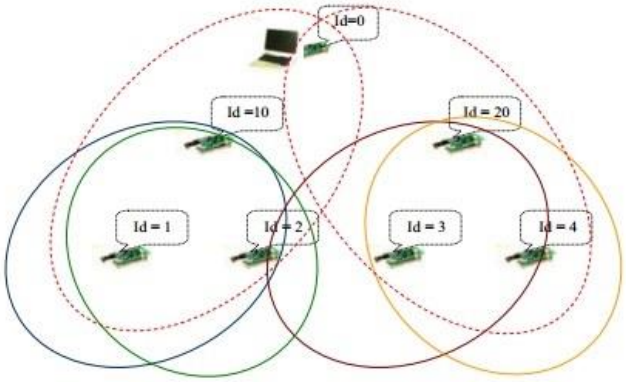

Figure 1. Multi-hop data Collection (LWSDR) [3]

The second set of data is collected from eKo Pro Series sensor nodes commissioned with two eS1100 soil moisture sensor experimented at Network Laboratory, Sikkim Manipal Institute of Technology (SMIT) as shown in the figures below (Figure 2, Figure 3).

Figure 2. Soil Moisture Data being analyzed on EkoPro Series Motes

Copyright ⓒ 2015 SERSC

The spatially and temporally associated soil moisture data is collected from two closely monitored flower pots installed with soil moisture sensors as shown in Figure 3. At regular interval water is poured into flower pot to vary soil moisture readings.

Figure 3. Soil Moisture Sensors Installed on Flower Pots

The sensor node data from LWSDR and ekoView Data is converted into array data structure for programming purpose. The sample size is set to 20 which indicate that the program shall take 20 samples per analysis. At the end of simulation, the total number of samples that were similar based on similarity metrics (Pearson coefficient, Euclidian Distance & Cosine similarity) is calculated.

The snapshot of temperature data for outdoor sensor nodes - node 1 & 2 and temperature data for indoor sensor nodes – node 3 & 4 is given below in Table 2.

 Copyright ⓒ 2015 SERSC

Table 2. Snapshot of Temperature Data for Sensor Nodes

Reading #	Node 1	Node 2	Node 3	Node 4
1	30.21	30.16	27.61	27.63
2	30.2	30.17	27.61	27.63
3	30.19	30.17	27.61	27.63
4	30.19	30.19	27.63	27.64
5	30.19	30.18	27.63	27.65
6	30.19	30.18	27.64	27.65
7	30.19	30.18	27.64	27.66
8	30.19	30.17	27.65	27.66
9	30.21	30.17	27.65	27.67
10	30.22	30.18	27.66	27.69
11	30.23	30.18	27.66	27.69
12	30.23	30.18	27.67	27.69
13	30.23	30.18	27.67	27.7
14	30.24	30.17	27.67	27.71
15	30.24	30.18	27.67	27.71
16	30.23	30.18	27.67	27.73
17	30.23	30.18	27.67	27.73
18	30.23	30.18	27.67	27.73
19	30.22	30.19	27.68	27.73
20	30.21	30.19	27.69	27.73

6. Results

The program discussed in Section 4 is run with experimental temperature and humidity data obtained from Labeled Wireless Sensor Network Data Repository (LWSDR). The percentage of energy conservation in indoor and outdoor sensor from LWSDR data source is discussed below in Table 3.

Table 3. Experiment Result (for LWSDR data source)

Parameters	Similarity Metrics		
	Euclidean Score	Cosine Similarity	Pearson Coefficient
Temperature	2.95 %	14.95 %	14.95 %
Humidity	0 %	6 %	6 %

Similarly, the percentage of energy conservation for soil moisture data obtained from Network Laboratory, Sikkim Manipal Institute of Technology (SMIT), Sikkim Manipal University is highlighted below in Table 4.

Copyright ⓒ 2015 SERSC

Table 4. Experiment Result (for SMIT data source)

Parameters	Similarity Metrics		
	Euclidean Score	Cosine Similarity	Pearson Coefficient
Soil Moisture	2 %	12.3 %	12.3 %

The percentage of energy conservation is directly proportional to total number of samples for which redundancy is observed as for those number of samples radio is switched to low energy consumption mode (like sleep mode). The formula for calculating energy conservation percentage from total number of identified redundant samples is given below:

$$Energy\ Conservation\ \% = \frac{0.01 \times Number\ of\ redundant\ \ samples}{Total\ samples}$$

As given is Table 3, one can conclude that 2.95 % of temperature data are similar as per Euclidean score but Cosine similarity and Pearson coefficient suggest that there is 14.95 % of similar temperature data. Also Cosine similarity and Pearson coefficient suggests 6 % of similar humidity data. For humidity, Euclidean score suggest no similarity at all.

As given in Table 4, one can conclude that Euclidean Score shows only 2 % similarity whereas Cosine similarity and Pearson Coefficient suggests 12.3 % of similar soil moisture data.

7. Discussion

The percentage of energy conservation discussed in Table 3 and 4 indicates that out of 3 similarity metrics, Pearson correlation coefficient and Cosine similarity gives sufficiently close result whereas Euclidean score is least reliable as per the experimental results. The result obtained in this paper seconds the observation made in [1] which suggests that Pearson correlation coefficient gives better result than any other method. The overall energy consumption in wireless sensor network can be calculated by following formula [8]

Overall Energy Consumption = Energy for sensing + Energy for localization + Energy for data processing + Energy for transmission (transmission, receiving) – Energy depletion

The model provided in this paper exploits energy consumption by reducing unneeded communication and sensing. Let *E(i)* be total energy consumption without applying the technique presented in this paper and *E(j)* be total energy consumption with the technique mentioned in this paper. *E(i) > E(j*) as its shown experimentally in Section 6 that by exploiting similarity metrics one can reduce unneeded communication and sensing as presented here in terms of energy conservation percentage. Overall Energy Consumption is inversely proportional to Energy Conservation %. *i.e.*, more is energy conservation %, less will be overall energy consumption thereby increasing overall lifetime of WSN.

8. Conclusion

As sensor node deployments become dense, the sensed data becomes closely associated. The model presented in this paper identifies the set of sensor nodes which captures and communicates similar data taking the aid of similarity metrics like Euclidean score, Pearson coefficient and Cosine similarity. The similarity metrics can be used to identify redundant

 Copyright ⓒ 2015 SERSC

nodes with the help of an algorithm running at the base station. The sensor nodes with redundant samples can be switched to low energy consumption mode for energy conservation purposes. The experiment suggests that application specific configuration is required in order to propose efficient data-aware energy conservation schemes based on similarity metrics. Mission critical applications may not benefit from this approach as loosing actual data from sensor nodes may create erroneous results. It is also observed that Pearson correlation coefficient and Cosine Similarity gives sufficiently reliable result than Euclidean Score. The future work that can be done in this regard is to apply data forecasting algorithm to conserve energy even more.

References

[1] I. F. Akyildiz, Ian F., Weilian Su, Yogesh Sankarasubramaniam, and Erdal Cayirci. "Wireless sensor networks: a survey." Computer networks 38, no. 4 **(2002)**: 393-422.
[2] B. Q. Ali, N. Pissinou and K. Makki, "Identification and Validation of Spatio-Temporal Associations in Wireless Sensor Networks", Sensor Technologies and Applications, 2009. SENSORCOMM'09. Third International Conference, IEEE, **(2009)**, pp. 496-501.
[3] G. Anastasi, "Energy conservation in wireless sensor networks: A survey", Ad hoc networks, vol. 7, no. 3, **(2009)**, pp 537-568.
[4] J. Balendonck, J. Hemming, B. A. J. Van Tuijl, L. Incrocci, A. Pardossi and P. Marzialetti, "Sensors and wireless sensor networks for irrigation management under deficit conditions (FLOW-AID)", Proceedings of the International Conference on Agricultural Engineering (AgEng 2008), **(2008)**.
[5] W. R. Heinzelman, A. Chandrakasan and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks", System sciences, 2000. Proceedings of the 33rd annual Hawaii international conference, IEEE, **(2000)**, pp. 10.
[6] Y. Jiber, H. Harroud and A. Karmouch, "Precision agriculture monitoring framework based on WSN", Wireless Communications and Mobile Computing Conference (IWCMC), 2011 7th International, IEEE, **(2011)**, pp. 2015-2020.
[7] R. Jurdak, A. G. Ruzzelli and G. MP O'Hare, "Radio sleep mode optimization in wireless sensor networks", Mobile Computing, IEEE Transactions, vol. 9, no. 7, **(2010)**, pp. 955-968.
[8] E. S. Nadimi, V. Blanes-Vidal, R. Nyholm Jørgensen and S. Christensen, "Energy generation for an ad hoc wireless sensor network-based monitoring system using animal head movement", Computers and Electronics in Agriculture, vol. 75, no. 2, **(2011)**, pp. 238-242.
[9] S. Suthaharan, "Labelled data collection for anomaly detection in wireless sensor networks", Intelligent sensors, sensor networks and information processing (ISSNIP), 2010 sixth international conference on. IEEE, **(2010)**.
[10] M. C. Vuran, O. B. Akan and I. F. Akyildiz, "Spatio-temporal correlation: theory and applications for wireless sensor networks", Computer Networks, vol. 45, no. 3, **(2004)**, pp. 245-259.
[11] W. Dargie and C. Poellabauer, "Motivation for a Network of Wireless Sensor Nodes", Chapter 1 in book Fundamentals of Wireless Sensor Networks, Wiley, **(2010)**, pp. 3-14.
[12] W. Ye, Wei, J. Heidemann and D. Estrin, "An energy-efficient MAC protocol for wireless sensor networks", INFOCOM 2002. Twenty-First Annual Joint Conferences of the IEEE Computer and Communications Societies. Proceedings. IEEE, vol. 3, **(2002)**, pp. 1567-1576.

Copyright ⓒ 2015 SERSC

 Copyright ⓒ 2015 SERSC