A Data Prediction Model Based On Extended Cosine Distance For Maximizing Network Lifetime of WSN

ARUN AGARWAL* and AMITA DEV^
Department of Computer Science
Guru Gobind Singh Indraprastha University
Sector 16C, Dwarka, Delhi
India
Email: *arun.261986@gmail.com, ^amita_dev2@hotmail.com

Abstract: Wireless Sensor Network is a randomly deployed collection of sensor nodes with the aim to collect the information near its sensing area. The approach is to sense some event, record its respective data value and transmit it to a sink where this data value is utilized and thus become information. The values received by the sink may contain duplicate, inappropriate and inconsistent values. The new research design may focus on collecting and sending only that value which may be utilized by the sink. The transmission of irrelevant data is avoided to increase the performance and lifetime. This paper aims at providing a data management technique which reduces load on sensor nodes to enhance network lifetime. To reduce extra burden on sensor nodes a data prediction model is built which restricts data transmission by predicting future data values. The algorithm finds relationship between data values. The goal is to calculate degree of relatedness between these values so to establish a relation which predicts future value. The proposed approach is compared with existing linear regression model, dual prediction model and least mean square model. The result reflects that the approach presents better results in terms of prediction accuracy and total energy consumption.

Keywords: Clustering, Data Aggregation, Data Management, Data Prediction, Regression Analysis, Energy Efficiency

1. Introduction

Recent advances in communication and technology lead to many new inventions. Wireless Sensor Network is a classical example which fuses the advance wireless communication principles with the vast capacity of sensor nodes. WSN are designed to operate in isolated environment with the aim to collect data. The data collected by the sensor node is transmitted to base station. The major issue associated with WSN is limited energy. Thus it is required to control and manage the dissipation of energy in WSN to enhance lifetime. Most of the work carried out till date focuses upon reducing communication cost to improve performance. But WSN depends upon many other factors and data is one among them. Reducing or minimizing amount of data for transmission and aggregation has a great impact on performance of WSN [1, 2].

WSN may be used in variety of applications such as health monitoring, smart homes, environment monitoring, military applications etc. Environment monitoring is the most widely used application of WSN but it comes with inherent limitation of limited power and it usually requires a long network lifetime to obtain desired results [3]. The network lifetime should be large enough so that it must collect a huge volume of data set which may be utilized to obtain some convincing results. Also the need is to continuously sense the environment for some sudden change. Various approaches have been given in literature to reduce energy dissipation and every approach is different in terms of the factor considered for implementation. Some protocols may adopt efficient routing strategy; some may manage the amount of data transmitted. Data is the primary source of communication and the radio energy communication model is dependent upon the distance and total number of bits transmitted. Reducing data at each step of wireless sensor communication will result in improved result. Reducing the transmission cost based upon the reduction in total data transmitted will have better results in terms of prolonged network lifetime. Also there is negligible impact of increasing payload size so we may increase packet size.

This paper proposes a prediction model which is utilized to reduce total data transmission. The prediction model measures the amount of relatedness between data values to build the prediction model. The variation in distance is analyzed and the prediction model is applied to predict future values. Thus rather than transmitting actual data values, values are predicted by the base station itself. The overall transmission time is divided into two halves one active mode and other one is idle mode. In active modes nodes are operational which means they are directed to keep themselves active and perform the desired task of data sensing, transmission, aggregation and receiving. Whereas in idle mode sensor nodes keep them in non operational or sleep mode. In this time period nodes switch off and will use minimum energy. This approach results in reduced overhead and increased performance. This paper uses a large data set which is collected over a long duration, and prediction model is build based upon this large data set thus to encompasses a wide domain of data values and to attain high prediction accuracy. The proposed approach is a two step process where first step is to build the prediction model and second step is to use this model to predict future values. In the first step the nodes are in active mode and send all their sensed data to base station. Base station collects this data and build prediction model. The proposed prediction model gives a variation to existing cosine distance. Cosine distance is commonly used in several data management applications to measure relatedness between data values.

The remainder of this paper is organised as follows. Starting with Section 2 which gives a brief study of various papers and contribution related to data prediction techniques. This section discusses various approaches that may be used to manage the transmission of data. Section 3 describes the system model and proposed approach. It gives the equations and prerequisites to build proposed system model. Section 4 gives the details of system parameters and analyses the performance of proposed protocol in terms of energy reduction and model fitting value. Comparison with similar approaches has been given in this section. Section 5 concludes the study and also lists limitations and future aspects of this study.

2. Literature Survey

A wide variety of work is carried out in field of data management using data prediction technique. The aim is to reduce data which will result in reduced transmission cost. Many papers have been published and each uses different approach to reduce data and improve performance. Somasekhar et al. [4] proposed a pre filtration method. This paper presents a correlated method in which relations between data transmissions are identified. Samer Samarah[5] proposes a prediction model based upon integration of WSN and cloud computing. This approach builds a prediction model and data is directly disseminated at cloud for future predictions. Using cloud for storing sensed data values reduces storage cost at sensor node.

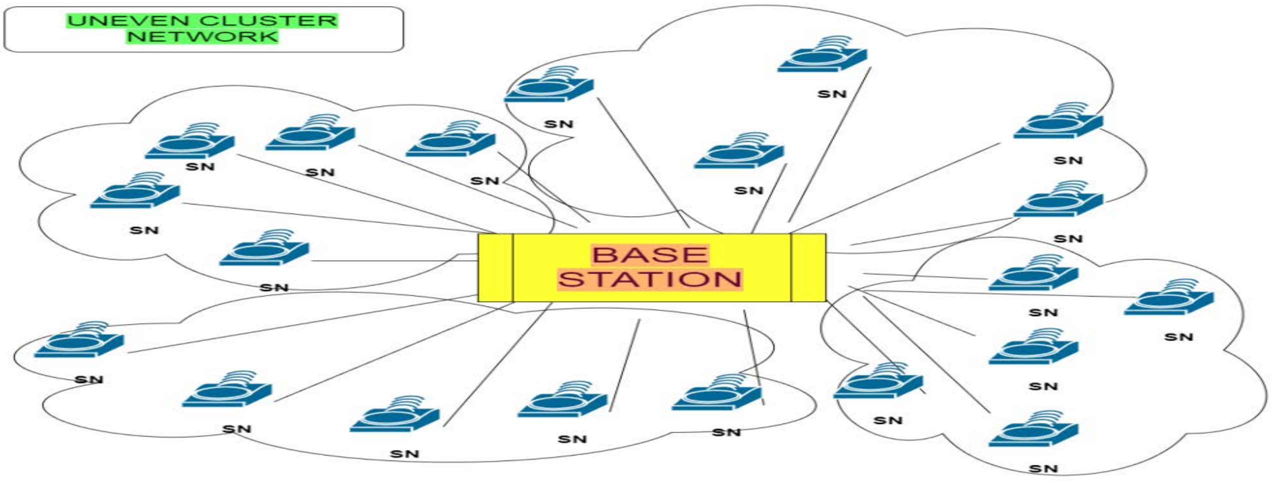

Fig.1: A Portion of Uneven Clustered Network with 20 SN

This integration of cloud with sensor nodes leads to new dimensions in WSN research area.

Mou et al. [6] in their paper monitors several environmental parameters and it combines data prediction with compression. The application of least mean square method for prediction is used where CHs obtain an approximated value after stipulated time period. Data prediction in WSN guesses new values based on the old values received in past. There exist some models in literature which enhances performance by adopting data prediction mechanism. Auto regressive model uses linear regression and this approach has been used by Tulone et al. [7]. Guiyi et al [8] presents a different approach to reduce data by removing temporal redundancy. The authors proposed a double queue mechanism to predict data values and to maintain synchronisation between communicating nodes. This paper uses grey model and Kalman filter for data aggregation and integrate both these filters to achieve better performance. Wang et al. [9] suggested a method to remove geographical redundancy. Correlations between data values are established.

Several other methods have been proposed to reduce total communication cost. This paper presents comparison with the existing simple regression technique, dual prediction mechanism and least mean square algorithm. An alternative is proposed to enhance performance and to prolong network lifetime.

3. Proposed Approach

This paper presents a regression model which establishes relation between data values. The network is composed of two layered architecture, where lower layer is composed of sensor nodes and upper layer contains base station. This paper considers a portion of network consisting of 20 sensor nodes which directly communicate with base station as shown in figure 1.

The transmission of data from sensor nodes to base station is carried out in terms of direct communication, where all the sensor nodes transmit their sensed data directly to base station. The first order radio communication model is used. The objective of the proposed approach is to acquire complete information without transmitting total data. The proposed system uses a technique to reduce total number of transmissions. Sensor nodes are directed to operate in two modes active and idle. In active mode each sensor node senses the environment, collect the data value, prepare a message and transmit the same to base station. Base station will collect all these values for a fixed duration $\tau = \{t_1+t_2+t_3........+t_n\}$, where time duration is divided into equal time slots and sensor node will generate one data value per time slot. All the sensor nodes [1: N], will collect their sensed data values and transmit an ordered pair including the data value d_j and corresponding node id j.

i.e. $\forall$ j = [1: N], send $m_j(d_j,j)$ to BS

Base station will collect all this data into a two dimensional vector D of size [τ, N], where τ is total active time divided into equal time slots and N is the total number of sensor nodes. Now the base station will construct the following data vector at the end of time duration τ, given in Table 1.

Time\ NodeID	SN1	SN2	SN3	....	N^{th} SN
t1	**R**	**R**	**R**	**.....**	**R**
t2	**R**	**R**	**R**	**.....**	**R**
....	**.**	**..**	**..**	**.....**	**..**
t_n	**R**	**R**	**R**	**.....**	**R**

Table 1: Collection of data values by base station

By using the collected data values over a given time period the base station will construct the prediction model by using the following terms:

$$\text{Å}[\mathbf{1}:\boldsymbol{\tau}] = (\textstyle\sum_{\mathbf{1}}^{\mathbf{N}} \mathbf{Rj})/\mathbf{N} \ , \quad \text{---- (1)}$$

For all nodes j= [1: N], it gives the average data value per time slot for each sensor node for given duration τ.

$$\hat{\text{H}}[\mathbf{1}:\boldsymbol{\tau}] = \mathbf{max}(\mathbf{Rj}) \ , \quad \text{---- (2)}$$

For all nodes j= [1: N], it gives the maximum data value per tie slot for each sensor node for given duration τ.

$$\hat{\text{h}}[\mathbf{1}:\boldsymbol{\tau}] = \mathbf{min}(\mathbf{Rj}) \ , \quad \text{---- (3)}$$

For all nodes j= [1: N], it gives the minimum data value per time slot for each sensor node for given duration τ.

Using the above vectors in equation 1, 2 and 3 the proposed algorithm calculates the weight of the prediction which is further used to build prediction model.

$$\boldsymbol{\alpha} = \sum_{\mathbf{1}}^{\boldsymbol{\tau}} \sqrt{\frac{(\hat{\mathrm{H}} - \hat{\mathrm{h}})^{\mathbf{2}}}{\text{Å}}}, \quad \text{----} \ (4)$$

$\boldsymbol{\alpha}$ is the prediction weight for time duration τ.

Using the data given in table 1, i.e. the time duration τ, number of sensor node values N and corresponding data reading we will calculate the fitting factor to normalize the predicted values by applying the following equation:

$$\boldsymbol{\beta} \ = \frac{(\sum\sum \boldsymbol{Rij})}{\boldsymbol{N*t}}, \quad \text{----} \ (5)$$

$\boldsymbol{\beta}$ is the fitting factor for time duration τ.

Now the prediction equation takes up the form

$$\check{Y} = (\boldsymbol{\alpha\,\tau} + \boldsymbol{\beta}),$$

$\check{Y}$ are calculated predicted values for next τ time.

In next time slot the sensor nodes are directed to keep them in idle mode to save energy. During this sleep period, base station will predict the data values based upon the proposed Regression Model. The sensor node will save its balance energy during this time period. This will enhance the overall lifetime to a greater instant. The analysis is carried out to determine total amount of energy saved during this period.

4. Simulation And Results

The proposed model is applied on data set where the values are collected for very large duration. The simulation uses total 100 sensor nodes and random network architecture is assumed to implement scenario that closely resembles to real situations. The randomness of the network is obtained by assigning random positions to all the sensor nodes in the beginning of each round. Various environment parameters have been collected and the same equation is applied to predict different data set values. Initially temperature is taken to analyze our prediction model. For simulation purpose a large data set is used which consists of thousands of data values per sensor node. The graph shown below in figure 2 and 3 gives a snapshot of complete simulation where the data values collected over a given time period is aggregated and summarized by 20 sensor nodes. All the data values are collected during active lifetime of sensor node. When sensor nodes switches itself in sleep mode base station will predict data values and use the same as if this data value is received by the respective sensor node. The difference between predicted value and actual value is calculated and results have been depicted in figure 2.

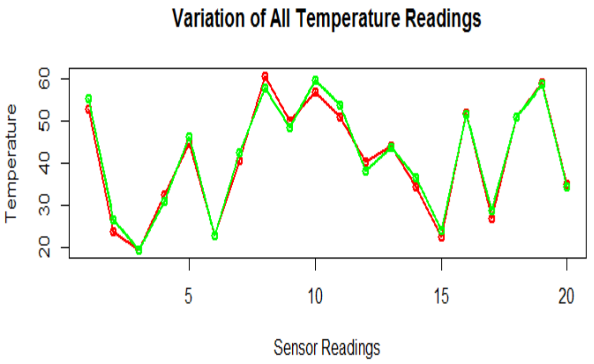

Fig.2: Actual vs. Predicted Value for 20 SN

It is clear from the above figure that predicted data – red color, coincides with the actual data – green color, i.e. the difference between actual data and predicted data is very less. This small difference shows that these values are highly correlated and instead of transmitting actual data values base station will substitute these values by the predicted values. By data prediction sensor nodes will save its scarce energy source by keeping them in idle mode during half of their lifetime. In other words we may say that these predictions will double the overall network lifetime as compared with data transmission without prediction.

The prediction error is calculated in terms of root mean square value which is depicted in above figure 3.

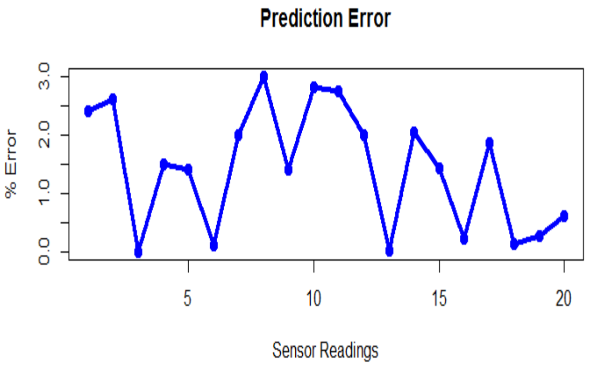

Fig.3: Prediction RMS error for 20 SN

From the above figure 3 it is clear that root mean square lies within acceptable range of permitted errors. For some values RMS error is almost zero and for other values it ranges from 0.0 to 3.0. This gives us a quantitative measure to find degree of relatedness between data values.

The proposed model is compared with simple regression model, dual prediction model and least mean square model. The performance is compared in terms of prediction accuracy and proportion of energy reduction. The results have been listed in table 2.

Prediction Model	Prediction Accuracy %	Proportion of Energy Reduction
Simple Regression	82.60	0.9275
Dual Prediction	86.75	0.8750
Least Mean Square	87.50	0.8285
Proposed Model	94.95	0.7960

Table 2: Result Analysis in terms of prediction accuracy and energy reduction

Simulation is being carried out several times to test the randomness of network. Each time the performance is measured in terms of residuals which are the difference of actual value to the predicted value. The residuals are used to calculate prediction accuracy and the results have been compared with other prediction approaches. Also the paper calculates proportion of energy reduction which is the ratio of energy consumed using prediction to the actual energy consumed without prediction. Both the results are tabulated above.

5 Conclusion

In this paper, we proposed a new framework for processing environmental data in WSN, which uses data prediction. In our prediction model the distance between data values is calculated. This distance gives the relationship of degree of relatedness between data values. The measure of correlation is given by the proposed prediction model equation. The proposed equation is formulated in terms of prediction weight and fitting factor. The attractive feature of this approach is that it reduces total data communication. The nodes are allowed to transmit only for fixed time duration than it is switched off and will be kept in idle mode to save energy. The results have been analyzed and tabulated above. Based upon the results obtained we conclude that our proposed approach is better in terms of prediction accuracy. Also there is a significant reduction in total energy dissipation which enhances overall performance of the network.

The proposed method when compared with other techniques will give far better performance. In our next papers we conduct our research to calculate the actual time values for active and sleep time. The impact of varying sleep time may be analyzed and also it may be interesting to test the performance of the network in case where sleep time is more than active time period.

References:

[1] I.F. Akyildiz, M.C. Vuran: Wireless Sensor Networks, *In: John Wiley & Sons*, 2010.

[2] G. Wener-Allen, K. Lorincz, M. Ruiz, O. Marcillo, J. Johnson, J. Lees, M. Walsh: Deploying a wireless sensor network on an active volcano, Data-Driven Applications in Sensor Networks, *In: IEEE Internet Computing*, March/April 2006.

[3] Kulik, Heinzelman And Balakrishnan: Negotiation-Based Protocols for Disseminating Information in Wireless Sensor Networks, *In: Wireless Networks 8*, 169–185, 2002.

[4] Somasekhar Kandukuri, Jean Lebreton, Richard Lorion, Nour Murad, and Jean Daniel Lan-Sun-Luk: Energy-Efficient Data Aggregation Techniques for Exploiting Spatio-Temporal Correlations in Wireless Sensor Networks, *In: IEEE Transactions* 2016.

[5] Samer Samarah: Data Predication Model for Integrating Wireless Sensor Networks and Cloud Computing, *In: Procedia Computer Science 52* (2015) 1141 – 1146.

[6] MouWu, Liansheng Tan, Naixue Xiong: Data prediction, compression, and recovery in clustered wireless sensor networks for environmental monitoring applications, *In: Information Sciences 329* (2016) 800–818.

[7] D. Tulone, S. Madden: Time series forecasting for approximate query

answering in sensor networks, *In: Proceedings of the 3rd European Conference on Wireless Sensor Networks (EWSN)*, 2006, pp. 21–37.

[8] Guiyi Wei , Yun Ling a, Binfeng Guo, Bin Xiao, Athanasios V. Vasilakos: Prediction-based data aggregation in wireless sensor networks - Combining grey model and Kalman Filter, *In: Computer Communications 34* (2011) 793–802.

[9] Haiying Shen, Ze Li, Lei Yu, Chenxi Qiu: Efficient Data Collection for Large-Scale Mobile Monitoring Applications, *In: IEEE Transactions On Parallel And Distributed Systems*, Vol. 25, No. 6, June 2014.

[10] Subir Halder, Amrita Ghosal: A survey on mobility-assisted localization techniques in wireless sensor networks, *In: Journal of Network and Computer Applications 60*(2016)82–94.

[11] K. Bicakci, I.E. Bagci, B. Tavli: Communication /computation tradeoffs for prolonging network lifetime in wireless sensor networks, *In: The case of digital signatures, Inf. Sci. 188* (2012) 44–63.

[12]Guo, W., Xiong, N., Vasilakos, A. V., Chen, G., & Cheng, H: Multi-source temporal data aggregation in wireless sensor networks, *In: Wireless Personal Communications*, (2011), 56, 359–370.

[13] Sinha, A., & Lobiyal, D. K.: Probabilistic data aggregation in information-based clustered sensor network, *In: Wireless Personal Communications*, (2014), 77(2), 1287–1310.

[14] Edara, P., Limaye, A., & Ramamritham K: Asynchronous in-network prediction- Efficient aggregation in sensor networks, *In: ACM Transactions on Sensor Networks*, (2008), 4(4), 25–34.

[15] H. Jiang, S. Jin, C. Wang: Prediction or not? An energy-efficient framework for clustering-based data collection in wireless sensor networks, *In: IEEE Trans.Parallel Distrib. Syst. 22* (6) (2011) 1064–1071.

[16] Zhang Z., Deng B., Chen S., Li L: An Improved HMM Model for Sensing Data Predicting in WSN. *In: Web-Age Information Management. WAIM 2016. Lecture Notes in Computer Science*, vol. 9658. Springer, Cham.

[17] Guo, W., Xiong, N., Vasilakos, A. V., Chen, G., & Cheng, H: Multi-source temporal data aggregation in wireless sensor networks, *In: Wireless Personal Communications*, (2011), 56, 359–370.