

# Data Reduction in Wireless Sensor Network: A Survey

<sup>1</sup>Manisha P. Mashere, <sup>2</sup>Sunita S. Barve, <sup>3</sup>Pramod D. Ganjewar

<sup>1,2,3</sup>Dept. of Computer Engineering., MIT Academy of Engineering, Pune, Maharashtra, India

## Abstract

Wireless sensor network is deployed in remote and hostile areas where no infrastructure is available. Wireless sensor network includes sensor nodes for sensing to monitor physical and environment condition. Recently, Wireless Sensor Network (WSN) used in many areas like military, environment monitoring, hospitals, biological equipments, biomedical, health monitoring etc. The limitation of WSN includes lifetime of network, Battery, Bandwidth, Energy etc. In this paper, we mainly focused on data reduction for Energy minimization in the network. Data reduction is one of the data pre-processing techniques of data mining that can increase storage efficiency and reduce costs. Data reduction (DR) aims to remove frequently occurring data while transmission. For this purpose many data reduction strategies are introduced depending on the scenario of WSN. Recent data reduction algorithms and techniques are introduced in this survey, which helps to increase energy as well as prolong the lifetime of the network.

## Keywords

WSN, Data reduction, Energy efficiency, Network Lifetime

## I. Introduction

WSN is an emerging technology that has a wide range of potential applications including environment monitoring, surveillance, medical systems, robotic exploration, military etc. The individual nodes in a wireless sensor network (WSN) are inherently resource constrained: they have limited processing speed, storage capacity, and communication bandwidth. After the sensor nodes are deployed, they are responsible for self-organizing an appropriate network infrastructure often with multi-hop communication with them. Then the on-board sensors start collecting information of interest.

WSN consist of a large number of distributed nodes that organized themselves into many multi-hop wireless networks. Each node equipped with one or more sensors, embedded processor and low power radios and is normally battery operated. A sensor node might vary in sizes and its cost. Generally, sensor node is a typical device that includes a sensing (for data acquisition from the environment), processing (for local data processing and storage) and communication (for data transmission).

In preceding existence, Wireless Sensor Networks (WSNs) have gained an amplified attention from their research community and extended its boundaries in commercial, industrial and medical domains. Sensor nodes equipped with one or more sensors, memory, processor, power supply, radio and an actuator. Verities of sensors can be attached further in network to measure properties of environment. Sensor has limited power supply or low power device which can't easily replace.

Data mining is a broad area contains data pre-processing techniques like data cleaning, data smoothing, data reduction, data compression etc. Other methods like classification, clustering and association mining mechanism also give attention in data pre-processing. Data reduction schemes used when samples are

unnneeded or null values is sensed by sensors and unnecessary for transmission. In-network processing is necessary when data aggregation occurs in between the sensor and sink. In this way the amount of data is reduced while traversing the network towards the base station.

## II. Related Work

WSN utilizes energy for data sensing, computing, and transmitting. Maximum Energy is required for data transmission from source node to sink node. In WSN consumed minimum energy using different ways like data compression, routing, clustering, data reduction, etc. Here data reduction techniques are explained.

## III. Data Reduction Techniques

### A. Data Prediction

Data prediction means to predict some value with the use of various algorithms and then choose the data accordingly. There are various types of algorithms which help for data reduction by prediction. For Example LMS (Least Mean Square) algorithm mostly used because of its simplicity and less complexity. Data prediction techniques can be divided into three categories (1) Stochastic approach [24]: It determines depiction of random process such that probabilistic model can predict the sensed values. Ken [25] uses dynamic probabilistic model and reduce communication overhead at sensor nodes.

Time-Series forecasting: These models are simple, but they are mostly used in practical cases because of good accuracy. PAQ [20] uses auto regressive models and predict the result of every single sensor. SAF (Similarity based Adaptive Framework) [19] works with general linear time series model which includes a time varying mechanism. It is used to identify outliers and inconsistent data. Moving Average (MA), Auto Regressive Moving Average (ARMA) etc. are Simple models and can used in practical cases. These methods give good accuracy and their implementation is simple and it's a lightweight approach.

Algorithmic approaches: It is used to get predictions, starting from heuristic or behavioral character of sensed values. For example PREMON [21] take snapshots of network by monitoring operation continuously. It's mostly suitable for cluster based WSN. Energy Efficient Data Collection (EEDC) [26] is good in inquiry-based applications, in which each sensor node relate to upper and lower bound and differences between bounds denotes the accuracy of sensed values. TGA [27] are used to improve coverage of WSN. Other prediction techniques are also useful in real time wirelessly sensor networks like LMS [10], WMA, ARIMA, Naïve Prediction model, PAQ [22] etc. with accuracy measures like Mean Absolute Error, Root Mean Square Error and Mean Absolute Percentage Error [23].

### B. Adaptive Filter

Filter is a device that reduces unnecessary features from a signal. Filter has two types stationary and non-stationary. Stationary filter doesn't allow variation of its component and non-stationary filters are required to track time variation and allow changing its

coefficient. Adaptive filter works as a same manner. It works with algorithms like LMS (Least Mean Square) [6], Kalman Filter [28], LMS-VSS( LMS Variable Step-Size) [29], LMS-SSA, Extended Kalman Filter [30] etc. for giving fast prediction and better result by changing its coefficients. LMS-SSA (LMS Step Size Adaption) gives more appropriate result in case of filtering in WSN and doesn't require any prior knowledge about correlation functions and matrix inversions and requires few computations. It is suitable for tiny sensor nodes which mostly used in WSN environment. Prediction [11] in WSN provides less transmission for sensor nodes by adjusting its filter parameters. Data will be transmitted with some prediction values and compare with another values of data counter and if there is mismatching, then alert message issued by sink and both ends switched to initialization mode otherwise standalone mode [31].

Adaptive filter also works with Dual Prediction framework where sensor and sink both has prediction model. It predicts local reading of sensor node and communication between nodes and sink will be reduced. Prediction can be done in both time and space through a pre-defined model whose parameters depends on historical data or a prior knowledge. LMS was implemented on FPGA (Field Programmable Gate Array) [28] to reduce the communication between the sensor nodes and base station.

### C. Voronoi Tessellation Transformation

The Voronoi Tessellation method applied to multimedia input data. This transformation used to reduce data and contents are Confidential [2]. It has a very low complexity. Also a novel integrated approach used to reduce the size of transmitting multimedia data.

### D. Dimension Reduction Method

Dimension Reduction method removes irrelevant, noisy and the redundant feature. Dimension reduction method is classified into FS (Feature Selection) and FE (Feature Extraction). This method reduces dimension in feature space and improve performance of the clustering algorithm [4]. The first FS method proposed is three stages FS-FS-FE. FS method obtains the relevant feature, the second FS method removes the redundant feature and FE method reduces the dimension. This method provides a low dimensional feature space. Dimension reduction method is effective to remove irrelevant, redundant and noisy features and preserving valuable information in the high dimensional data sets.

### E. Compressive sensing-based EEG (electroencephalogram) and Transmission EEG Feature

In this paper present energy efficient data reduction approaches for reducing transmission data in a wireless EEG seizure detection [5]. In seizure detection system, reduce data using compressive EEG sensing based method. This method, transmitting the EEG signal is to compress the raw data before their transmission. The EEG compression algorithm has been proposed in [32]. In the transmission of EEG feature method transmits only section or feature of the signal in seizure detection [5]

### F. Adaptive and Efficient Dimension Reduction Model (APCADR)

In hierarchical sensor networks consist of adaptive and efficient dimension reduction model. (APCADR) proposed for hierarchical sensor networks based on the candid covariance-free incremental

PCA (CCIPCA) [7]. PCA is applied for dimensionality reductions in WSN were proposed in [33, 34]. A PCA based data compression model proposed in [15].

A Principal component analysis (PCA) is a well-known multivariate data analysis technique. This technique used for reduced dimensionality. PCA is used for calculation of the covariance matrix. The basic PCA algorithm is based on the calculation of the covariance computational complexity.

### G. Fuzzy Logic

Lotfi A. Zadeh has proposed the concept of fuzzy logic in 1965 [9]. Fuzzy logic is a multi-valued logic. It formalizes reasoning when deal with vague terms. In which decisions are not limited to either true or false, or as with Boolean logic 0 or 1. Therefore Fuzzy logic algorithm takes into consideration the degree of truthfulness and falsehoods. Crisp set of input data is converted into a fuzzy set using fuzzy linguistic variables (representation of the system's input or output variables), terms and membership functions (used to map non fuzzy input values) this process called fuzzification. Now some rules based approaches applied and defuzzification step process denotes the output by its membership functions. Fuzzy logic is helpful to send only the decision of sensor nodes towards the sink node generally used for humidity, temperature, light intensity etc. It is used as a solution of power consumption in WSN via reducing messages which are sent to the sink.

### IV. Proposed Work

Sampling techniques definitely give us better results. In addition to the sampling techniques CDC can also add its own benefit with a sampling technique to prolong the life of network by reducing the transmission data.

### V. Conclusion

The main goal of this survey paper is used to minimize energy consumption through the data reduction method in WSN and thus increasing the network lifetime. This survey paper gives an overview of recent update in data reduction technique. Many of the articles cited in this paper give their contribution and real time application as well as helpful for future work. The data reduction approach is based on Wireless Sensor Network topology. Some of stated algorithms provide good results. Different types of algorithm provide many techniques which can be combined to overcome the individual limitations and measures of system performance.

### References

- [1] Muruganantham Arunraja, Veluchamy Malathi, Erulappan Sakthivel, "Energy conservation in WSN through multilevel data reduction scheme", *Microprocessors and Microsystems* Vol. 39, pp. 348–357, 2015.
- [2] A. Mostefaoui, H. Noura, Z. Fawaz, "An integrated multimedia data reduction and content Confidentiality approach for limited networked devices", *Ad-Hoc Networks*, Vol. 32, pp. 81–97, 2015
- [3] Taner Cevik, "Energy Conservative Multitier Architecture with Data Reduction for Cluster-Based Wireless Sensor Networks", *International Journal of Distributed Sensor Networks* Vol. 2015, pp. 11, 2015.
- [4] Kusum Kumari Bharti, P.K. Singh, "A three-stage unsupervised dimension reduction method for text clustering", *Journal of Computational Science*, Vol. 5, pp. 156–169, 2014.
- [5] Joyce Chiang, Raba K. Ward, "Energy-Efficient Data Reduction Techniques for Wireless Seizure Detection

- Systems", *Sensors open access*, Vol. 14, pp. 2036-2051, 2014.
- [6] S. Haykin, "Least-Mean-Square Adaptive Filters", New York Wiley, 2003
  - [7] Murad A. Rassam, Anazida Zainal, Mohd Aizaini Maarof, "An adaptive and efficient dimension reduction model for multivariate wireless sensor networks applications", *Applied Soft Computing*, Vol. 13, pp. 1978-1996, 2013.
  - [8] M.Arun raja, V. Malathi, "An LMS Based Data Reduction Technique for Energy Conservation in Wireless Sensor Network (WSN)", *Int.Journal Computer Technology & Applications*, Vol. 3, pp. 1569-1576, July-August 2012.
  - [9] Mohamed O. Abdel-Aal, Rabie A. Ramadan, "Energy Saving and Reliable Data Reduction Techniques for Single and Multi-Modal WSNs", Vol. 978, No. 1, pp. 4673-4810, 2012.
  - [10] Silvia Santini, Kay Romero, "An adaptive strategy for quality-based data reduction in wireless sensor networks, in: *Proceedings of the 3rd International Conference on Networked Sensing Systems (INSS06)* 2008.
  - [11] Pham, Ngoc Duy, Trong Duc Le, Hyunseung Choo, Enhance exploring temporal correlation for data collection in WSNs, in: *IEEE International Conference Research, Innovation and Vision for the Future*, IEEE, 2008.
  - [12] I. K. Fodor, A Survey of Dimension Reduction Techniques, Lawrence Livermore National Laboratory, US Department of Energy, 2002.
  - [13] I.T. Jolliffe, *Principal Component Analysis*, 2nd ed., Springer, New York, 2002.
  - [14] Y. Le Borgne, G. Bontempi, Unsupervised and supervised compression with principal component analysis in wireless sensor networks, in: *Proceedings of the Workshop on Knowledge Discovery from Data, 13th ACM International Conference on Knowledge Discovery and Data Mining*, pp. 94-103, 2007.
  - [15] W.K.G. Seah, Z. A. Eu, H. Tan, "Wireless sensor networks powered by ambient energy harvesting (WSN-HEAP) - Survey and challenges, *Wireless Communication, Vehicular Technology, Information Theory and Aerospace & Electronics Systems Technology, Wireless VITAE 2009. 1st International Conf. on Vol.1, No. 5, pp. 17-20, May 2009.*
  - [16] B. Kanagal, A. Deshpande, "Online Filtering, Smoothing and Probabilistic Modeling of Streaming Data", *Proc. 24th International Conf. on Data Engineering (ICDE 2008)*, Cancún, México, April 7-12, 2008.
  - [17] S. Santini, K. Romero, "An Adaptive Strategy for Quality-Based Data Reduction in Wireless Sensor Networks," *3rd International Conf. on Networked Sensing Systems (INSS 2006)*. TRF, pp. 29-36, Chicago, IL, USA, June 2006.
  - [18] N. Paul, C. Debono, "An FPGA implementation of adaptive data reduction technique for WSN," *Proc. WICT 08 conference*, Dec. 2008.
  - [19] T. Ross, "Fuzzy logic with engineering applications," John Wiley & Sons Ltd, 2004
  - [20] D. Tulone, S. Madden, "Time series forecasting for approximate query answering in sensor networks", In: *proc., 3rd European Conference on Wireless Sensor Networks (EWSN06)*, February 21-27, 2006.
  - [21] Y. Le-Borgne, S. Santini, G. Bontempi, "Adaptive model selection for time series prediction in wireless sensor networks", *Signal Process*, Vol. 87, No. 12, pp. 3010-3020, 2007.
  - [22] D. Tulone, S. Madden, "Time series forecasting for approximate query answering in sensor networks", *Third European Conference on Wireless Sensor Networks*, pp. 21-37, 2006.
  - [23] Femi A. Aderohunmu, Giacomo Paci, Davide Brunelli, Jeremiah D. Deng, Luca Benini, Martin Purvis, "An application specific forecasting algorithm for extending WSN lifetime", *IEEE International Conference on Distributed Computing in Sensor Systems*, 2007.
  - [24] D. Chu, A. Deshpande, J.M. Heller Stein, W. Hong, "Approximate data collection in sensor networks using probabilistic models", In: *proc. 22nd International Conference on Data Engineering (ICDE06)*, Atlanta, GA, Vol. 48, pp. 3-8, 2006.
  - [25] D. Chu, A. Deshpande, J.M. Heller stein, W. Hong, "Approximate data collection in sensor networks using probabilistic models", *22nd International Conference on Data Engineering 2006*, pp. 48-59, 2006.
  - [26] Q. Han, S. Mehrotra, N. Venkata subramanian, "Energy efficient data collection in distributed sensor environments", in: *Proceedings of the IEEE International Conference in Distributed Computing Systems*, pp. 590-597, 2004.
  - [27] Shibo He, Jiming Chen, Xu Li, Xuemin (Sherman) Shen, "Leveraging Prediction to Improve the Coverage of Wireless Sensor Networks", *IEEE Transactions on Parallel and Distributed Systems*, Vol. 23, No. 4, April 2012.
  - [28] Nicholas Paul Borg, Dr. Carls James Debono, "An FPGA Implementation of an Adaptive Data reduction Technique for WSN", Dept. of communication and computer engineering, University of Malta.
  - [29] Biljana Stojkoska, dimitar Solev, Danco Davce, "Data Prediction in WSN using variable step size LMS Algorithm". *SENSORCOMM 2011: 5th International Conference on Sensor Technologies and Applications*.
  - [30] L. Ramrath, A. Schlaefer, F. Ernst, S. Dieterich, A. Schweikard, "Prediction of respiratory motion with a multi frequency based Extended Kalman Filter", *21st International Conference and Exhibition on Computer Assisted Radiology and Surgery (CARS'07)*, Vol. 21, 2007.
  - [31] R. Kwong, E. W. Johnson, "A variable step size LMS algorithm", *IEEE Trans. On signal process.* Vol. 40, pp. 1633-1642, 1992
  - [32] Antoniol, G. Tonella, P., "EEG data compression techniques", *IEEE Trans. Biomed. Eng.* Vol. 44, pp. 105-114. 1997.
  - [33] Y. Le Borgne, G. Bontempi, "Unsupervised and supervised compression with principal component analysis in wireless sensor networks", *Proceedings of the Workshop on Knowledge Discovery from Data, 13th ACM International Conference on Knowledge Discovery and Data Mining*, pp. 94-103, 2007
  - [34] Y.A. Le Borgne, S. Raybaud, G. Bontempi, Distributed principal component analysis For wireless sensor networks, *Sensors*, Vol. 8, pp. 4821-4850, 2008.