Prediction Based Compression Algorithm for Univariate and Multivariate Data in Wireless Communication Networks

N. Vini Antony Grace[1], A. Chilambuchelvan[1], J. Shanmugapriyan[2]

[1] Department of Electronics and Instrumentation Engineering, R.M.D. Engineering College, India
[2] Department of Electrical and Electronics Engineering, Vel Tech Multi Tech Dr. Rangrajan Dr. Sakunthala Engineering college, India
antonygrace81@gmail.com, chill97@gmail.com, jspriyan@gmail.com

Abstract

In wireless sensor network, energy is primary constraint in designing and installation. WSN network possess many sensor nodes and each are equipped with batteries as its power bank. Due to the usage of many sensor nodes and batteries, there comes need to reduce the power consumption and increase the efficiency. The data sensed can be compressed before transmission to improve efficiency and reduce power ingestion. The data sensed through WSN network are time series data. Data prediction plays a crucial role in forecasting data with past values, which reduce the usage of sensor nodes in sensing information. This paper proposes the differencing based prediction algorithm and leveling based compression algorithm for both univariate and multivariate data. This system reduces the number of sources and power consumption. The algorithms were simulated using a simulation tool MATLAB 8.2- R2013a. The algorithms are quantified by calculating approximation error, compression ratio, and computation complexity on the compressed data. From the results it is evident that the proposed algorithm is better suitable for compression of univariate and multivariate signals.

Keywords: Approximation mean error, Compression ratio, Energy consumption, Multivariate data

1 Introduction

Large scale wireless sensor networks (WSNs) are employed in real-time monitoring and surveillance applications, such as monitoring the environment [1] monitoring health care of remote [2], classification of animal behavior [3] and structural health monitoring for infrastructures [4]. More data at high speed will help the application to perform efficiently and effectively. The same can be realized by more sensor sampling and this will lead to large volume of raw data from sensors. Consequently, the sensor data requires more energy for transmitting and huge memory for storage. That is the battery supplying energy to the sensor for acquiring data at remote location will discharge at a high rate for transmitting all the sampled data to the monitoring station. As per Kimura and Latifi [5] each remote sensor consumes 80% of battery capacity for data transmission. The results presented by Barr and Asanovic [6] provides the evidence that the energy consumed by the network for sending a single bit of information is almost same as the energy required by the processor for performing thousands of operations. Also more energy consumption lead to more heating of sensor nodes and thus reduce the life of battery. Thus it becomes a big challenge to operate, maintain and manage the data acquisition and storage systems in application scenarios.

Among several approaches adopted to optimize energy requirements and improve the life of sensor, battery replacement, sample rate reduction and data compression [7] are most adopted in the practical scenario. In practical and physical situations battery replacement will be expensive and impossible. It is practically impossible to change batteries in sensor nodes for many applications because they are mostly used in harsh and unreachable environments. Thus the existence of sensor nodes depends strongly on battery lifetime. So to increase the network lifespan of WSN, the energy consumption of each sensor node sub unit has to be cautiously handled.

Reduction of sampling rate can moderate the data volume and memory requirements but will lead to loss of required information. Lower sampling rate can shrink the data volume for transmitting and storing, but the reduced data often loose informative details of the monitored system. A higher sampling rate will always be preferred to achieve quality data for better system performance and will be also helpful in arriving decisions.

On the other hand, data compression proves to be the more promising technique that provides good data quality, best of system performance and considerably reduces the energy consumption of WSNs. In literature many algorithms and techniques have been presented

*Corresponding Author: N. Vini Antony Grace; E-mail: antonygrace81@gmail.com
DOI: 10.3966/160792642020032102017

and their capability to compress the time series data is proven and established. However many of them are derived and tested for compression of single variate sensor data, for example temperature, humidity, etc. In addition WSNs are implemented to acquire multi-dimensional data such as to measure acceleration and strain during seismic in three axis (x, y and z), radiated emission data from an electronic equipment during EMI test, signals sensed by magnetometer to identify magnetic field's direction and the EEG (Electroencephalography) signals providing details about the behaviour of heart or brain. In spite of vast applications of multivariate signals only few literature works are available addressing the issues in systems used to acquire, analyse and store multivariate signals.

Organization of the Paper

The organization of this paper is as follows: In Section 2, a brief about the related work is given. The prediction algorithm, univariate and multivariate lasso compression algorithm and the methodology of implementation is proposed in Section 3. Section 4 briefs the various evaluation parameters used for assessing the performance of the proposed algorithm. In section 5 the description of real world publicly available data sets is given and the results of the proposed algorithm are compared with several compression algorithms available in literature. Finally, section 6 briefs few concluding remarks.

2 Related Work

A wireless sensor network (WSN) is a network composed of sensor nodes communicating among themselves and deployed in large scale (from tens to thousands) for applications such as environmental, habitat and structural monitoring, disaster management, equipment diagnostic, alarm detection and target classification [8].

A simple routing protocol algorithm was proposed by [9]. The sender sensor can transmit the information by sending joining request to the intermediate nodes, which has the stability to pass the information. By this algorithm one can find the optimal route to transfer the information to the destination node.

The paper proposed by [10] gives stability equation to find the stable sensor nodes near sender nodes. Due to this we can increase the life time of the sensor nodes, save the energy on sensors for transmitting. However, the maximum energy used by sensor node is in sensing and transmitting the large amount of information sensed.

The Energy consumption on the WSN network can be reduced by employing the sensors by calculation region of coverage or Region of interest [11]. The Region of coverage is the region which comes under the sensing unit of each sensor.

A triangular scheme algorithm has been developed by [10] for employing every sensors out of the coverage areas of their neighbor sensor nodes. The triangular scheme helps in finding the coverage areas without leaving holes between the sensor nodes. This also reduces travelling distance of the sensor nodes. It may increase the utilization of sensing area but not effective collection of information.

The main ways to reduce energy consumption is prediction and compression [12-13]. The prediction is the important topic in machine learning that the machine can forecast the data with its previous observed data. This plays a crucial role in energy conservation, source deficiency etc.

In paper [14] the compression has been done with PPM (prediction by partial matching) combine with Context Tree weighting for FSMX sources. These compression algorithms can be only suitable for source of binary data. Computational cost is high for this method and it is very difficult to implement this work. This method can give high efficiency in text compression.

There are many prediction based compression algorithms which uses DBN, ARIMA, Universal algorithms [15-17] for time series analysis. Most of the prediction based compression process is there for univariate data. The univariate data are the data that differs in one dimension with change in time series [8]. Multivariate data is that each data will have more than one value for each time instances. It differs in more than one dimension.

In this paper an efficient algorithm is proposed which performs prediction and compression on univariate and multivariate time series data. The prediction algorithm is included before compression to reduce the energy ingestion by sensor nodes, which gives better results in compression than the previous works. The experimental results of proposed system performed significantly better when tested with smooth multivariate data sets. This makes the system most suitable in applications such as behavior monitoring.

3 Proposed System

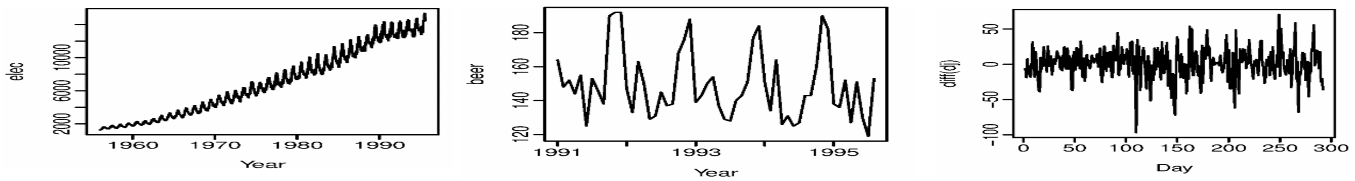
3.1 Prediction Algorithm

Prediction algorithm has made to adopt for time series analysis. It can be done with the previous dataset collected through the sensor nodes. With the past values, the future values can be forecasted and also the thousands of dataset values can be shrunk to hundreds before transmission. In receiving end the original set of values can be retrieved by using same prediction algorithm. Data prediction plays a crucial role in forecasting data with past values which reduce the usage of sensor nodes in sensing information and also reduces power consumption of nodes.

Since the values will change with the varying time, the first process in algorithm is to draw the time series plot to know the nature of dataset values taken for prediction. Then the data series are differenced to make it stationary. The stationary series is time invariant and its mean, variance and auto-covariance are constant so that dataset values can be estimated well.

The above Figure 1(a), Figure 1(b) and Figure 1(c) show the time series plot of different quantities. The difference in the consecutive observations in the original series can be written as

$$Y_t' = Y_t \sim Y_{t-1} \tag{1}$$



(a) Monthly electricity production

(b) Monthly beer production

(c) White noise plot of Dow jones stock market index

Figure 1.

Let y denotes the r^{th} difference of Y which denotes the set of past values

If r=0: $y_t = Y_t$

If r=1: $y_t = Y_t - Y_{t-1}$

If r=2: $y_t = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2})$

The series is differenced with the lag of 1. If the series is not stationarized with first order differencing, then second and third order differencing can be done to make it stationary.

The stationarized values show the random walk model with T-1 difference series. Then the autocorrelation function (ACF) and partial autocorrelation function (PACF) can be calculated to find the order ‘p’ of auto regressive process and ‘q’ of moving average process.

The mean of the time series $Y_t, Y_t, \ldots, Y_n$

$$\hat{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i \tag{2}$$

Auto-covariance function at lag k, for $k \geq 0$

$$S_k = \frac{1}{n}\sum_{i=1}^{n-k}(Y_i - \hat{Y})(Y_{i+k} - \hat{Y}) \tag{3}$$

Auto-correlation function at lag k for k>0

$$\rho = \frac{S_k}{S_o} \tag{4}$$

Partial Autocorrelation function measures the degree of association between the variable (Y) to be predicted and the other time variable ($X_1, X_2, \ldots, X_3$), which is used in predicting the Y.

$$PACF = \frac{\operatorname{cov}(Y, X_1 \mid X_2, X_3, X_4)}{\sqrt{\operatorname{var}(Y, X_2, X_3, X_4)\cdot \operatorname{var}(X_1, X_2, X_3, X_4)}} \tag{5}$$

PACF will identify the order ‘p’ of Auto regressive term and ACF will identify the order ‘q’ of Moving Average term. After finding the p and q, the value can be predicted by using the general predicting equation given below

$$\hat{Y}_t = 1 + \varphi_1 B + \varphi_2 B^2 + \cdots + \varphi_p B^p - \theta_1 B - \theta_2 B^2 + \cdots + \theta_p B^q \tag{6}$$

$$\varphi(B) Y_t = \theta(B) \in_t \tag{7}$$

$$\varphi(B) = 1 + \varphi_1 B + \varphi_2 B^2 - \cdots - \varphi_p B^p$$
$$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 - \cdots - \theta_p B^q$$
$$BY_t = X_{t-1} \tag{8}$$

B: Lag operator

$\varphi_i (i = 1, 2, \ldots, p)$ are autoregressive parameters

$\varphi_i (i = 1, 2, \ldots, q)$ are moving average parameters

$\varepsilon_t \sim (0, \sigma^2)$ is the error term that follows a normal distribution.

3.2 Compression Algorithm

After predicting the values with prediction algorithm, the data is checked whether it is univariate or multivariate data. The compression algorithm for both univariate and multivariate data are discussed in this section.

3.2.1 Univariate Data

The univariate data varies in one dimension with respect to the time changes. The examples of the univariate are seismic data of volcanoes, price of the chemical product changing over, relative humidity, temperature, and blood pressure.

Let the observed one dimensional reading be $X = (X_1, X_2, \ldots, X_M)$ and the corresponding time instances were $(t_1 \leq t_2 \leq \cdots \leq t_M)$. It can be seen that the values measured by y will be varying during every time instant in a given time interval t, every values of

X captured by the sensor in the time interval t should be acquired for monitoring and stored for further studies, analysis, modelling and development. On the other hand, it can be for a sequence of instances, the change in values of X can be of less significance.

In some cases, the noted reading has same values with slightest perturbations with change in the time period such as $X_a = x_{a+1} = \cdots X_b$ and the $a < b$. This data can be stored in the database by $X_a = (\cdots, X_a, X_{b+1}, \cdots)$ with corresponding time instances $t = (\cdots, t_a, t_{b+1}, \cdots)$. On finding the time range (t_a, t_b) of unchanged values, the storage space to hold these readings can be reduced by $|a-b|$ and its corresponding time instances. These values are not exactly constant but with slightest perturbation in the time range (t_a, t_b). To remove these minor perturbations, we provide leveling approximation to minimize the difference in the slightest perturbations

$$J(y) = \frac{1}{2}\sum_{m=1}^{M}(X_m - Y_m)^2 + \gamma\sum_{m=1}^{M-1}|Y_{m+1} - Y_m| \quad \textbf{(9)}$$

The first part of the equation $\frac{1}{2}\sum_{m=1}^{M}(X_m - Y_m)^2$ is to reduce the difference between noted reading x and approximation y. The degree of regularization γ is a constant to trade-off the scale between the approximated and actual value which can control the levelness of the approximation. The higher values of γ gives rise to better approximation and higher compression ratio.

3.2.2 Multivariate Data

Multivariate data comprises of data extending in one or two dimensions for example movement of our hands that gives us corresponding movement data in X, Y, Z directions like acceleration. Torque and Magnetic field orientation. More than one dimension data will come under the multivariate data.

Now the algorithm can be extended to compress univariate data to be suitable for multivariate data. The observed multivariate time series be $X \in R^{DXM}$ and the corresponding time instances with $(t_1 \le t_2 \le \cdots \le t_M)$. D corresponds to the dimensions of each multivariate datasets. There will D readings at each time instances. Then we can write the time series as $X_{d*} = X_{d1}, X_{d2}, \cdots, X_{dM}$ is the d-throw in the matrix X. We will get D readings $X_{1m}, X_{2m}, \cdots, X_{dm}$ at the m-th time instance t_m.

The straight forward formulation to make the univariate algorithm to fit for the multivariate signals is as below:

$$J(Y) = \sum_{d=1}^{D} J(Y_{d*}) \quad \textbf{(10)}$$

Where Y_{d*} is the d-th row in the solution matrix $X \in R^{DXM}$

$$J(Y_{d*}) = \frac{1}{2}\sum_{m=1}^{M}(X_{dm} - Y_{dm})^2 + \gamma\sum_{m=1}^{M-1}|Y_{d,m+1} - Y_{dm}| \quad \textbf{(11)}$$

Therefore it follows that

$$\begin{aligned} J(Y) &= \sum_{d=1}^{D} J(Y_d) \\ &= \frac{1}{2}\sum_{m=1}^{M}\| X_{*m} - Y_{*m} \|^2 + \gamma\sum_{m=1}^{M-1}|Y_{*,m+1} - Y_{*m}| \quad \textbf{(12)} \\ J(Y_{d*}) &= \frac{1}{2}\| X - Y \|_F^2 + \gamma\sum_{m=1}^{M-1}|Y_{*,m+1} - Y_{*m}| \end{aligned}$$

The above expression will be used in compressing the multivariate data by reducing perturbations along with piecewise linear constant. However the above expression may fail to compress the Y optimally. To provide the optimal solution it is better to provide the projection of multivariate data by introducing the objective function.

$$J(Y, V) = \frac{1}{2}\| XV - Y \|_F^2 + \gamma\sum_{m=1}^{M-1}|Y_{*,m+1} - Y_{*m}| \quad \textbf{(13)}$$

By mininmizing the function J(Y, V) with an optimum *V*, the energy efficiency of original multivariate can be realized where *V* is an orthogonal matrix such that $V^T V = I$ and I is an identity matrix.

3.3 Algorithm of the Proposed System

The algorithm of the proposed system is given in Table 1.

4 Performance Assessment Measures

4.1 Compression Ratio (CR)

The compression ratio is described as the ratio between size of the compressed signal Y to the size of the actual uncompressed signal from sensor X.

$$CR = 100 \times \left(1 - \frac{compressed\ size\,(Y)}{Uncommpressed\ size\,(X)}\right) \quad \textbf{(14)}$$

4.2 Energy Consumption (EC)

The total energy consumed by a sensor node is contributed by the energy utilized for data compression and the energy required for transmitting the data. The energy requirements for data compression is mainly dependent on the energy required for one cycle (E_{opt})

Table 1. Algorithm of the proposed system

Algorithm
A. Prediction
1. Take the previous dataset of the values to be predicted
2. Draw the time series plot for the given data
(a) Use difference equation $Y_t' = Y_t \sim Y_{t-1}$ to make the time series plot stationary
(b) Find the ACF and PACF to find order of Autoregressive and moving average process 'p' and 'q'.
(c) Predict the data using general equation
$\hat{Y}_t = 1 + \varphi_1 B + \varphi_2 B^2 + \cdots + \varphi_p B^p - \theta_1 B - \theta_2 B^2 + \cdots + \theta_p B^q$
B. Compression
1. If(data= univariate)
Compress data using leveling equation
$J(y) = \frac{1}{2}\sum_{m=1}^{M}(X_m - Y_m)^2 + \gamma\sum_{m=1}^{M-1} \lvert Y_{m+1} - Y_m \rvert$
2. Else(data = multivariate)
Compress data using equation
$J(Y_{d*}) = \frac{1}{2} \lVert X - Y \rVert_F^2 + \gamma\sum_{m=1}^{M-1} \lvert Y_{*,m+1} - Y_{*m} \rvert$
To optimally compress the multivariate data, minimize the function
$J(Y,V) = \frac{1}{2} \lVert XV - Y \rVert_F^2 + \gamma\sum_{m=1}^{M-1} \lvert Y_{*,m+1} - Y_{*m} \rvert$
C. Transmit the compressed data

and number of CPU cycles (M) required for compression and the requirements by its peripherals are considered negligible. Energy required for transmission and receiving the data depends on the energy consumed by equipment to transmit a bit of data (E_{bit}) and the number of bits (N). The metric for estimating the energy consumption is as follows:

$$E_{total} = M \times E_{opt} + N \times E_{bit} \quad \textbf{(15)}$$

4.3 Approximation Mean Error (AE)

The approximation error for each compression algorithm is computed using the following equation:

$$AE = \frac{1}{N} \lVert XV - Y \rVert_F^2 \quad \textbf{(16)}$$

Where X and Y represent the actual multivariate data and its compressed form. The transformation matrix $V = I$ is the identity matrix for univariate algorithms. For multivariate data compression, Equation 13 gives the choice of V.

5 Experimental Results and Discussion

The performance of algorithms on both multivariate and univariate datasets are illustrated and discussed in this section. The proposed algorithms are simulated on the platform MATLAB 8.2- R2013a and its output comparison graphs and tabulations are shown under this section.

5.1 Experimental Data

Sensor Scope in collaboration with HES-SO and Clim-Arbes project a small sensor network is deployed during summer on the border of Le Boiron De Morges. The dataset consists of univariate data providing the readings of temperature and humidity. It is found the range of the value in humidity is comparatively larger than temperature. It is also noticed that the data is smooth without any sharp variations. These two datasets are taken at the sampling interval of 2 minutes.

One more univariate data set is obtained by single Microphone, deploying sensor nodes at Strata Clara convention center1. The data collected can be considered as discontinuous or non-smooth since it has a lot of noises and surges. This data is taken at the sampling interval 4 to 9 seconds.

Besides, first multivariate dataset is obtained from the UCI machine learning [8] database collected by MHealth (mobile health) program. These data based on multi-modal sensing on different parts of a body comprises of human behavior analysis. The data records comprises of body motion and vital signs of physical activities like running, standing and bending etc. The action experienced by different body parts like acceleration, rate of return and magnetic field orientation were recorded by placing the sensor on the volunteer's chest, left ankle and right wrist. The data taken from MHealth program are two lead ECG, Three axis acceleration data of right wrist and left ankle and three axis magnet data. These three multivariate data were taken at the sampling interval of 0.02 seconds. The first multivariate dataset were taken on one complete full day.

The second multivariate data set contains the three axis acceleration readings of vehicles and is published by CRAWDAD. The data set is collected by the motion of different vehicles using the mobile phone of its drivers.

The third multivariate data set contains the signals generated using three dimensional accelerator and used to monitor the behaviour of cow. The data is used to detect whether the cows are in the estrus. This data is taken at the sampling interval of 0.1 second for 24 days continuously. These past data values are taken to the prediction and then for compression. The data sets used for analyzing performance of algorithms are presented in Table 2.

Table 2. Dataset used for analysis

S.No	Univariate	Multivariate
1	Temperature	ECG (2 axis)
2	Relative Humidity	3 axis acceleration of Right Wrist
3	Microphone Data	3 axis Acceleration of Left Ankle
4	-	3 axis Magnet Data
5	-	3 axis acceleration data of motion of vehicle
6	-	3 axis accelerator reading in detecting cow's estrus behavior

5.2 Compression Performance

5.2.1 Univariate Data Sets

The performance metrics such as compression ratio, approximation mean error of proposed prediction based compression algorithm on univariate data set was compared with standard algorithms and are presented in Table 3 and plotted in graphs (Figure 2 to Figure 4).

Table 3. Performance comparison table for univariate dataset

Methods	Approximation Mean Error			Compression Ratio(%)		
	Temperature	Relative Humidity	Microphone data	Temperature	Relative Humidity	Microphone data
ULasso	2.4	4.1	6.35	99	94	97
LTC	1.2	2.7	4.4	99	98	96
PLAMLIS	0.8	2.2	4.4	98	93	92
Proposed	1.0	2.5	4.35	99	98	97

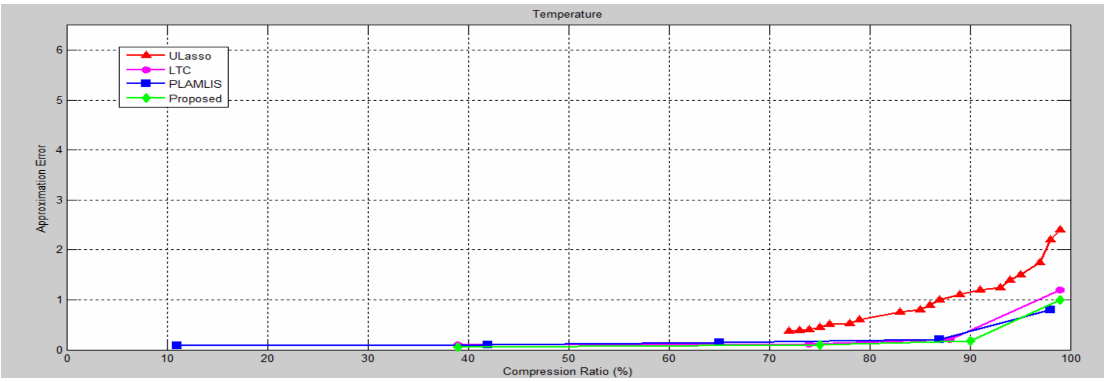


Figure 2. Compression ratio vs. approximation mean error for univariate data temperature

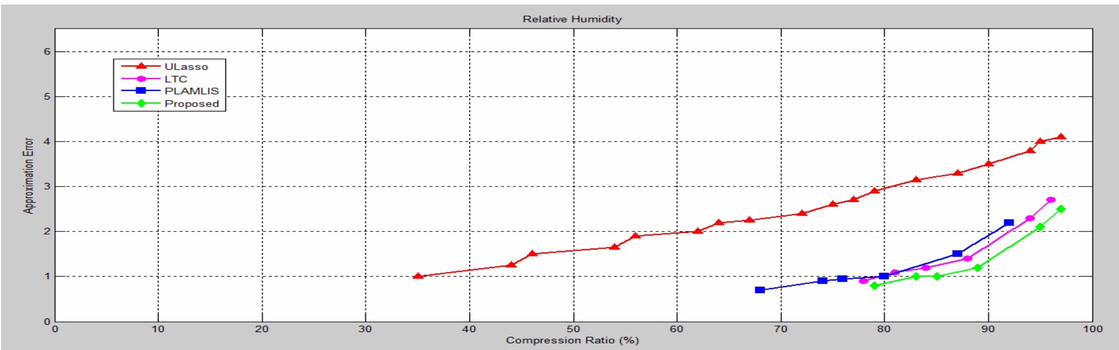


Figure 3. Compression ratio vs. approximation mean error for univariate data relative humidity

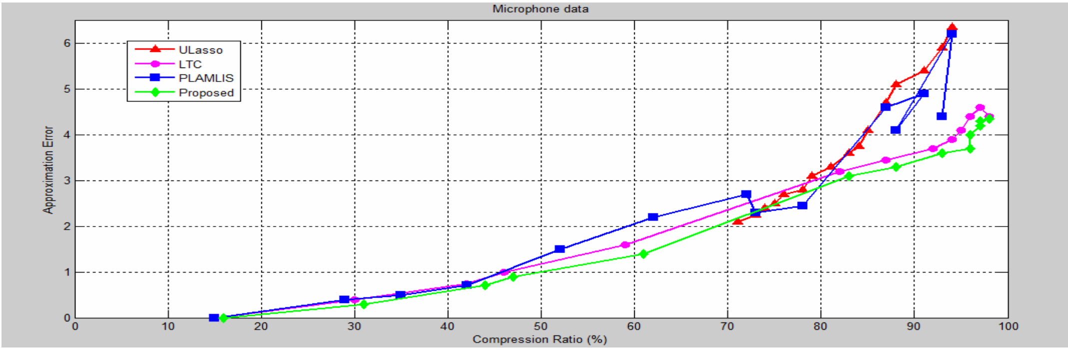


Figure 4. Compression ratio vs. approximation mean error for univariate microphone data

From the graphs it is inferred that, Approximation Error increase from lower to higher as the compression ratio increases. The proposed algorithm can produce less errors than standard algorithms for univariate data and it is shown that the efficiency of proposed algorithm is better.

5.1.2 Multivariate Data Sets

To compress multivariate data sets using algorithms like LTC and PLAMlis, first the multi-dimensional sensor signals are need to be divided into several univariate signals and each signal is compressed separately. The proposed algorithm operates considering the multivariate data completely as a whole and finds the next points in a multidimensional sequence.

The performance comparison of the proposed algorithm with conventional algorithms for multivariate data are plotted in graphs (Figure 5 to Figure 10). It is inferred that the Approximation Error of LTC and PLAMlis increases sharply with compression ratio, whereas for proposed algorithm, the Approximation Error is kept almost smooth and increasing slightly with respect to compression ratio. Moreover it can be seen that prediction algorithm compresses the signal with less Approximation Error. From the experimental results it shall be concluded that proposed algorithm is more suitable and produce better compression performance compared to conventional algorithms when implemented on smooth multivariate data signals and it performs as par to other algorithms on implantation to compress non-smooth multivariate data signals. The reason for better performance of proposed algorithm for multivariate signal is that it compresses the multivariate data as a whole. While LTC and PLAMlis that can produce better results on individual smooth data signals but the when synthesized will bring more error accounts in higher Average mean error.

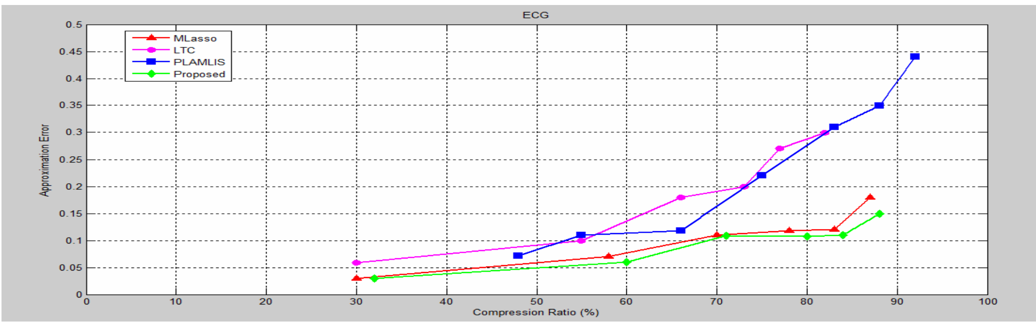


Figure 5. Compression ratio vs. approximation mean error for multivariate ECG data

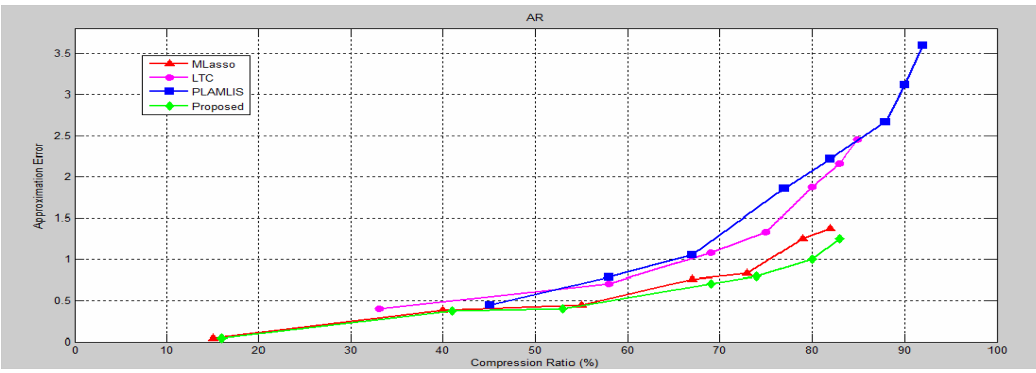


Figure 6. Compression ratio vs. approximation mean error for multivariate 3 axis data of right wrist

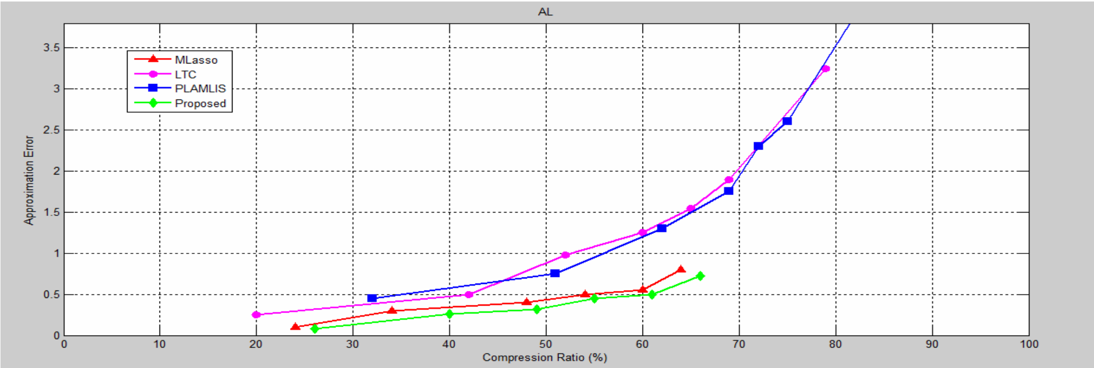


Figure 7. Compression ratio vs. approximation mean error for multivariate 3 axis data of left ankle

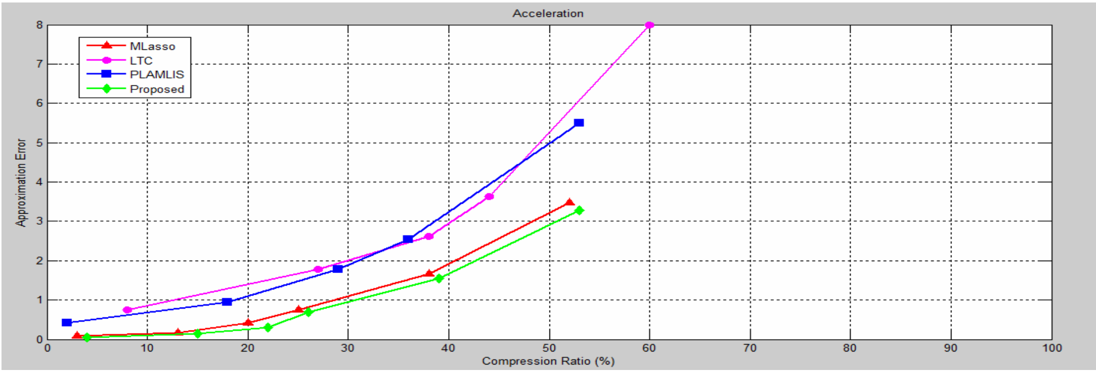


Figure 8. Compression ratio vs. approximation mean error for multivariate 3 axis magnet data

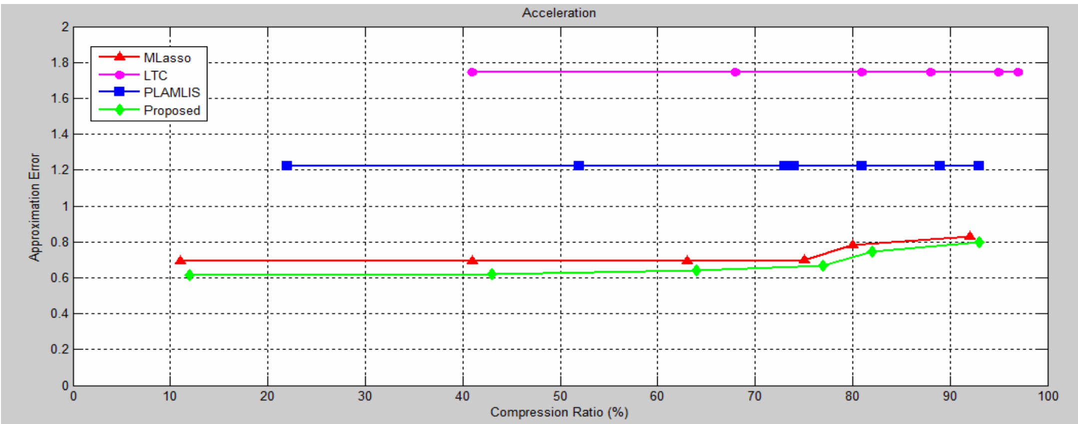


Figure 9. Compression ratio vs. approximation mean error for multivariate 3-axis accelaration readings of motion of vehicle

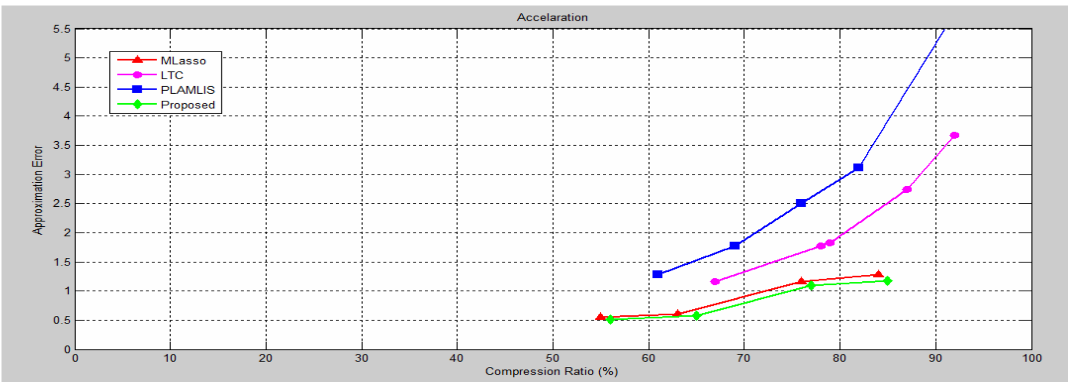


Figure 10. Compression ratio vs. approximation mean error for multivariate 3-axis accelarator reading on cow's estrus behaviour

It can be seen in the Table 4 and Table 5, the compression ratio of some conventional methods were higher than prediction based compression method. But they have higher mean error with higher compression ratio that could lead to the risk of losing important information in the data. Because of this reason, the proposed gives reasonable reduced error value with efficient compression ratio and better energy saving.

Table 4. Approximation mean error table of Multivariate datasets

Methods	Approximation Mean Error					
	ECG	Acceleration (Right Wrist)	Acceleration (Right Ankle)	Magnet Data	Acceleration (motion of vehicle)	Accelaration (cow Estrus Behaviour)
MLasso	0.18	1.38	0.8	3.466	0.83	1.281
LTC	0.3	2.45	3.25	8.00	1.748	3.67
pLAMLIS	0.44	3.6	4.28	5.568	1.22	5.50
Proposed	0.15	1.25	0.72	3.272	0.80	1.182

Table 5. Compression ratio table of multivariate datasets5.3. energy consumption

Methods	Compression Ratio(%)					
	ECG	Acceleration (Right Wrist)	Acceleration (Right Ankle)	Magnet Data	Acceleration (motion of vehicle)	Accelaration (cow Estrus Behaviour)
MLasso	87	82	64	52	92	84
LTC	82	85	79	60	97	92
pLAMLIS	92	92	84	36	93	91
Proposed	88	83	66	53	93	85

The main intention of compressing a signal is to save the computational time and energy saving. But to implement an algorithm to compress a signal, the system could consume some energy. It should be less than the energy saved by compressing the signal.

Energy consumed by conventional algorithm and proposed calculated by using the equation (15). With the calculated values, the graph (Figure 11 to Figure 12) was drawn to show the proposed work's efficiency. The most of the energy is consumed on transmitting and sensing. Obsolete work consumes energy both on transmitting and sensing but our proposed work consumes only on transmitting. The computation energy by CPU may be higher for proposed one but it is negligible when considering energy consumed during sensing.

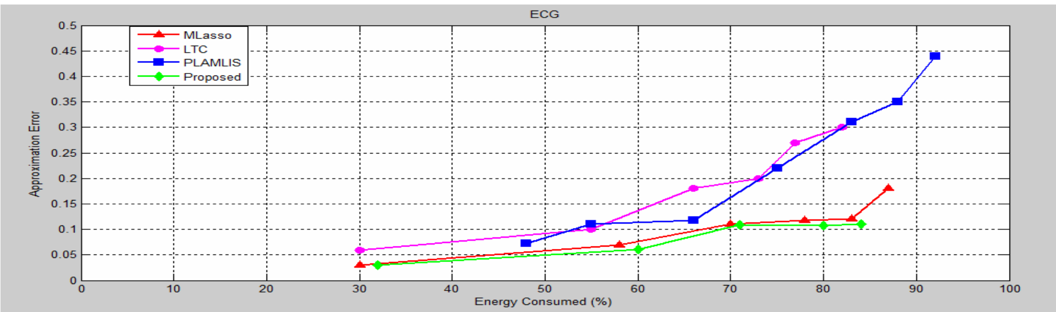


Figure 11. Energy consumption vs. approximation mean error for multivariate ECG data

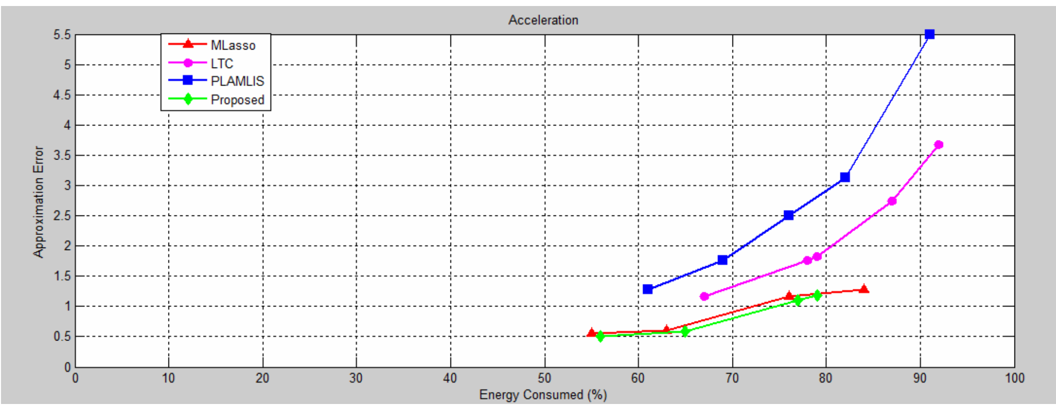


Figure 12. Energy consumption vs. approximation mean error for multivariate 3-axis accelarator reading on cow's estrus behaviour

The major disadvantage in LTC, PLAMLIS is that they change the multivariate data in to univariate data and then it will do the compression. Mlasso based algorithm operates considering the multivariate data completely as a whole and finds the next points in a multidimensional sequence which needs high computational time depending on size of the data scale [14] This may lead to additional need of memory space and high energy consumption. This disadvantage can be mitigated in our proposed work by forecasting values. Due to the inclusion of prediction algorithm, there is reduction in the energy consumption by sensor nodes. The proposed work saves 15% energy on ECG data prediction and compression. In Multivariate 3-axis accelerator reading on cow's estrus behavior data prediction based compression work saves 20% of energy. The prediction algorithm is also efficient in handling instabilities in the signals time series plot.

6 Conclusion

This paper proposes an efficient prediction and compression algorithm suitable for univariate and multivariate time series data as well. Instead of using computationally expensive transformations on raw data or introducing strong assumptions on data statistical models, the proposed prediction based compression algorithm forecast the future data for both stable and unstable data. So the proposed work gives lesser error than other works even though the datasets are differ in their statistical behavior. The performance of proposed algorithm is compared with conventional algorithms for three datasets of univariate and six datasets of multivariate category. As a result of comparison, the proposed algorithm gives highly and accurately forecasted data, maximum compression ratio up to 99% with less approximation mean error of 4.25 and reduced energy consumption by sensor nodes.

References

[1] L. Taheriazad, C. Portillo-Quintero, G. A. Sanchez-Azofeifa, Application of Wireless Sensor Networks (WSNs) to Oil Sands Environmental Monitoring, *Oil Sands Research and Information Network*, OSRIN Report No. TR-48, June, 2014.

[2] M. Kozlovszky, L. Kovacs, K. Karoczkai, Cardiovascular and Diabetes Focused Remote Patient Monitoring, *The Conference of VI Latin American Congress on Biomedical Engineering*, Paraná, Entre Ríos, Argentina, 2014, pp. 568-571.

[3] L. A. Gonzalez, G. J. Bishop-Hurley, R. N. Handcock, C. Crossman, Behavioral Classification of Data from Collars Containing Motion Sensors in Grazing Cattle, *Computers and Electronics in Agriculture*, Vol. 110, pp. 91-102, January, 2015.

[4] H.-N. Li, T.-H. Yi, L. Ren, D.-S. Li, L.-S. Huo, Reviews on Innovations and Applications in Structural Health Monitoring for Infrastructures, *Structural Monitoring and Maintenance*, Vol. 1, No. 1 pp. 1-45, March, 2014.

[5] N. Kimura, S. Latifi, A Survey on Data Compression in Wireless Sensor Networks, *International Conference on Information Technology: Coding and Computing*, Las Vegas, NV, USA, 2005, Vol. 2, pp. 8-13.

[6] K. C. Barr, K. Asanovic, Energy-Aware Lossless Data Compression, *ACM Transactions on Computer Systems*, Vol. 24, No. 3, pp. 250-291, August, 2006.

[7] E. Fasolo, M. Rossi, J. Widmer, M. Zorzi, In-Network Aggregation Techniques for Wireless Sensor Networks: A Survey, *IEEE Wireless Communications*, Vol. 14, No. 2, pp. 70-87, April, 2007.

[8] L. Yin, C. Liu, X. Lu, J. Chen, C. Liu, Efficient Compression Algorithm with Limited Resource for Continuous Surveillance, *KSII Transactions on Internet and Information Systems*, Vol. 10, No. 11, pp. 5476-5496, November, 2016.

[9] B. An, J. S. Lee, N.-S. Kim, D.-H. Kim, CARP: A Cooperative-Aided Routing Protocol in Mobile Ad-hoc Wireless Sensor Networks, *The 10th International Conference on Advanced Communication Technology*, Phoenix Park, Gangwon-Do, Korea, 2008, pp. 737-741.

[10] A. Reinhardt, D. Christin, M. Hollick, R. Steinmetz, On the Energy Efficiency of Lossless Data Compression in Wireless Sensor Networks, *The 4th IEEE International Workshop on Practical Issues in Building Sensor Network Applications*, Zürich, Switzerland, 2009, pp.873-880.

[11] R. J. D'Souza, S. Ganala, Optimal 2-Coverage by Heterogeneous Mobile Sensor Nodes Using Triangular Scheme, *Procedia Technology*, Vol. 4, pp. 187-195, June, 2012.

[12] M. A. Alsheikh, P. K. Poh, S. Lin, H.-P. Tan, D. Niyato, Efficient Data Compression with Error Bound Guarantee in Wireless Sensor Networks, *The 17th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, Montreal, Quebec, Canada, 2014, pp. 307-311.

[13] F. Marcelloni, M. Vecchio, An Efficient Lossless Compression Algorithm for Tiny Nodes of Monitoring Wireless Sensor Networks, *The Computer Journal Advance Access*, Vol. 52, No. 8, pp. 969-987, April, 2009.

[14] K. Sadakane T. Okazaki, H. Imai, Implementing the Context Tree Weighting Method for Text Compression, *Data Compression Conference*, Snowbird, Utah, USA, 2000, pp. 123-132.

[15] T. Hirata, T. Kuremoto, M. Obayashi, S. Mabu, K. Kobayash, Time Series Prediction using DBN and ARIMA, *International Conference on Computer Application Technologies*, Kunibiki Messe, Matsue, Japan, 2015, pp. 24-29.

[16] E. Federovsky, M. Feder, S. Weiss, *25th Annual International Symposium on Computer Architecture*, Barcelona, Spain, 1998, pp. 62-72.

[17] A. S. Lysyak, B. Y. Ryabko, Time Series Prediction Based on Data Compression Methods, *Problems of Information Transmission*, Vol. 52, No. 1, pp. 92-99, January, 2016.

Biographies

N. Vini Antony Grace is an Assistant professor in the department of EIE at R.M.D Engineering College. She is pursuing Ph.D. under Anna University in the area of Signal Processing. Her areas of interest include Digital Signal Processing, Wireless Communications. She has published 8 papers in National and International journals.

A. Chilambuchelvan is working as professor & Head in the department of EIE at R.M.D Engineering College. His areas of interest are Embedded system, VLSI design, soft computing. He has published 46 papers in International journals.

J. Shanmugapriyan is an Associate Professor in the department of EEE at Veltech Multitech Engineering College. His areas of interest are Power Electronics and Drives, and VLSI Design. He has published 10 papers in International and National journals.