Predicting and Modeling Biological functions in Body Area Network

Andrew Knox
Department of Electrical Engineering and Computing Systems
University of Cincinnati
Cincinnati, OH 45220-0008, USA
knoxar@gmail.com

Suryadip Chakraborty
, Department of Electrical Engineering, and Computing Systems
University of Cincinnati
, Cincinnati, OH 45220-0008, USA
suryadip.chakraborty@gmail.com

Dharma P Agrawal
Department of Electrical Engineering and Computing Systems
University of Cincinnati
, Cincinnati, OH 45220-0008, USA
dharmaagrawal@gmail.com

ABSTRACT
Recent advances in Wireless Body Area Sensor Network (WBASN) technology has become a leading approach for several promising applications in the medical field. WBASN is the network is built with different kinds of physiological sensors which measures different vital parameters on the human body for the purpose of monitoring the patients. Prediction and modeling are very two important issues which are required to address while building a WBASN. To enhance the long-term critical health monitoring, a robust predictive approach should be incorporated in every WBASN system and it leads to saving computation time and increasing the energy.

In this paper, we describe the use of polynomial regression for predicting and modeling biological functions. We also describe how effective different orders of polynomials can be. There are four functions that we use for this purpose: blood pressure, scalp EEG signals, the walking gait of people with neurodegenerative disorders, and lastly motor movement signals. We have also used two different degrees of polynomial functions to determine the predictive value.

Categories and Subject Descriptors
C.2.0 [Computer-Communication Networks]: General; D.0 [Software]: General; E.0 [Data] General; G.1.0 [Mathematics of Computing]: Error analysis;

General Terms
Design

Keywords
Wireless Body Area Sensor Network, Regression Polynomial, Data Aggregation, Correlation Coefficient, Prediction. Model

BODYNETS 2013, September 30-October 02, Boston, United States

DOI 10.4108/icst.bodynets.2013.253991

1. INTRODUCTION
Typically reading and recording biological functions requires a great deal of space. Recordings are often made at very small intervals and taken over a great deal of time requiring extremely large databases for this data. It also offers very little ability to make predictions for future readings. By using polynomial regression the only data that needs to be stored are the beta values of the polynomials. Then by using the time of day as input an approximate value for the given time can be acquired. Normally the advantages of this would be minimal; however, when used in conjunction with a wireless body area network (WBAN) the benefits are greater [1, 2].

A WBAN is a group a network computing devices that a user can wear on their body. Typically, this consists of several peripheral body sensors units (BSUs) and a single central unit to collect the data. The applications are numerous and include: monitoring vital signs of patients at risk of a heart attack, or detecting declines in insulin levels in diabetic patients. In addition, fact that the devices are wireless gives the patients freedom of movement that make them more attractive than traditional sensors. However, with this convenience comes a limitation, the greatest of which is providing power from a limited source. It is in this area the aggregation could provide improvements. Data aggregation can be implemented in a number of ways, such as clustering, data-centric routing, and others. Using polynomial regression for data aggregation provides much larger reductions in the amount of data that needs to be sent. In exchange the accuracy of the data is reduced. Where polynomial regression is most effective is when data is correlated over some dimension, in this case that is time.

2. BACKGROUND AND MOTIVATION
Data aggregation schemes in wireless networks tend to have one thing in common: preventing redundant data from being forwarded throughout the network [3]. To accomplish this, intermediate nodes are placed at junctions throughout the network. It is their job to collect data from lower nodes, analyze the data for redundancies, and pass on only the data with relevance and uniqueness. For example, wireless nodes that detect temperature in an outdoor environment are likely to have similar readings as those nearby. A simple aggregation scheme could be to simply pass only one temperature reading on to represent all the nodes connected to the sink. More complicated solutions are required for more complicated issues that arise.

If a network uses cluster heads, they are likely to drain their power source much faster since they are sending the data for many nodes rather than just their own information. One solution would be to shift cluster heads to power rich nodes in order to prolong the lifetime of the network [5].

A similar data aggregation scheme has been implemented in another work that is similar to what we propose though it is used for modeling temperature over a 2-d area. We felt that this same idea can be applied to body networks [4, 5].

We are using polynomial regression because the amount of data that would need to be sent by the sensors would remain stable over time, and is only dependent upon the order of the polynomial used. In addition, if a polynomial reasonably models the data, it can be used to predict data outside of scope of what has already been measured. This requires that the accuracy of the polynomial be high so that meaningful information can be predicted from it. Also, as WBANs become more and more common, it is likely the amount of data they will measure will increase, and the need to conserve power will become more and more necessary

3. CREATING POLYNOMIALS

Polynomial regression is a process of a relationship between a set of independent variables and a dependent variable is established, and modeled as an nth order polynomial. It is considered a special type of linear regression. In statistics it is used to provide an estimate of real data. It can also provide the ability to estimate values that are outside the range of original independent variables. Often it is displayed visually a line graph as shown in figure1.

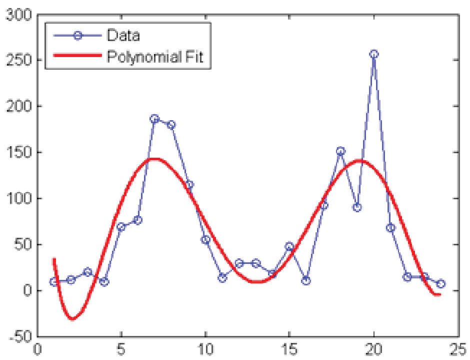


Fig1: Data representation by Polynomial Regression

3.1 Making Polynomials

For demonstration and comparison we used 4^{th} and 8^{th} order polynomials. Their polynomials have the form:

4^{th}:

$$f(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \beta_4 x^4$$

8^{th}:

$$f(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \beta_4 x^4 + \beta_5 x^5 + \beta_6 x^6 + \beta_7 x^7 + \beta_8 x^8$$

In this case the β values are the coefficients of the polynomial, and obtaining them is the primary goal of polynomial regression. If we start with a group of paired values (in this case a time and reading) and x is our time value, then y is our reading value. At this point two matrices need to be created in order to get the beta values we need. The following formulas are for a 4^{th} order polynomial. The 8^{th} order polynomial is similar in look though larger.

The first matrix:

$$A = \begin{pmatrix} n & \sum x_i & \sum x_i^2 & \sum x_i^3 & \sum x_i^4 \\ \sum x_i & \sum x_i^2 & \sum x_i^3 & \sum x_i^4 & \sum x_i^5 \\ \sum x_i^2 & \sum x_i^3 & \sum x_i^4 & \sum x_i^5 & \sum x_i^6 \\ \sum x_i^3 & \sum x_i^4 & \sum x_i^5 & \sum x_i^6 & \sum x_i^7 \\ \sum x_i^4 & \sum x_i^5 & \sum x_i^6 & \sum x_i^7 & \sum x_i^8 \end{pmatrix}$$

The second matrix: The beta matrix:

$$Z = \begin{pmatrix} \sum y_i \\ \sum x_i y_i \\ \sum x_i^2 y_i \\ \sum x_i^3 y_i \\ \sum x_i^4 y_i \end{pmatrix} \qquad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix}$$

The formula using these:

$A\beta = Z$, So, $\beta = A^{-1}Z$

3.2 Interpreting Results

To interpret the results, we assessed the results using the standard error, and the correlation coefficient.

3.2.1 Standard Error

$$std\ err = \sqrt{RSS/_{n-2}} \qquad RSS = \sum (y_i - \hat{y}_i)^2$$

Where $\hat{y}_i$ is the predicted value of y. Basically, RSS is the residual sum of squares.

3.2.2 Correlation Coefficient

$$corr\ coef = 1 - \frac{RSS}{SS_{total}} \qquad RSS = \sum (y_i - \hat{y}_i)^2$$

$SS_{total} = \sum (y_i - \bar{y})^2$; $\bar{y}$ = mean, and $\hat{y}_i$ =predicted value of y.

3.3 Limitations

Before Talking about the results of our work it is important to talk about the limitations of polynomial regression. First there is the fact that the polynomial is an *estimate* of the actual data. Therefore it is understood that it could not be used in situations where readings that are as accurate as possible are needed. In addition, biological signals are often in pulses, such as a heartbeat or the signals of a neuron. A polynomial is much better at modeling over trends of activity over time. Therefore, while it often does model the overall trend, it won't capture the movements on a small scale.

4. EXPERIMENTATION

Initially, data was obtained from PhysioBank ATM [6] which is website containing publically available data on a variety of different biological functions. The desired information can be downloaded as a text document that has columns for time, and reading. We wrote a program in Java that reads in the relevant file, performs polynomials regression on the data, and returns a text file with a third column containing the predicted results at each time stamp.

In addition, the standard error and correlation coefficient are calculated to provide a numerical measurement of the accuracy of the prediction. The Correlation coefficient is especially useful for this since it is easily interpreted. The standard error simply gives information on how much the prediction is likely to be wrong. A higher correlation coefficient is better, and a lower standard error is better. If the standard error was 0, and the correlation coefficient was 1.0, the polynomial would perfectly model the data (though this is an unlikely outcome for real world data).

This output file can then be imported into excel to provide visualization in the form of the graphs included in this paper.

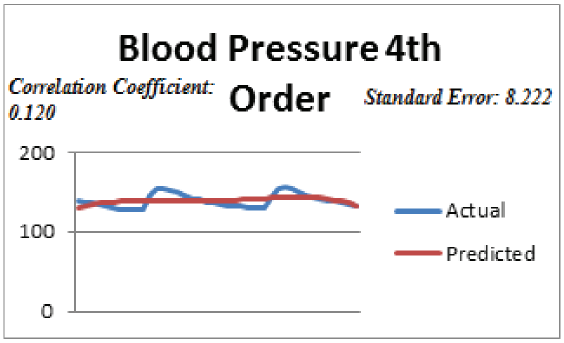
5. RESULTS

Below are the results from three different biological signals captured by a polynomial of order 4, and order 8 for comparison.

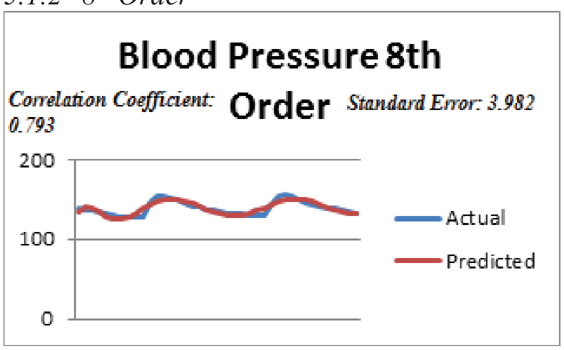
5.1 Blood Pressure

This is from a sample size of 80 where readings are taken at 0.01 second intervals. The readings are mmHg values.

5.1.1 4th Order



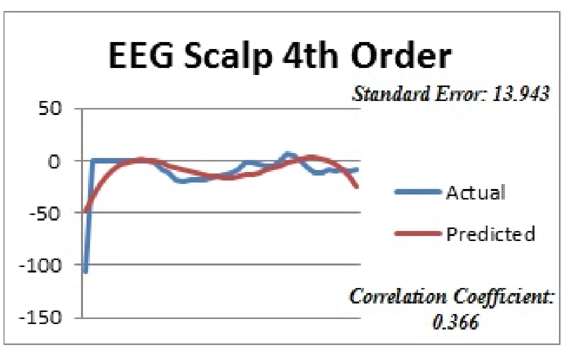
5.1.2 8th Order



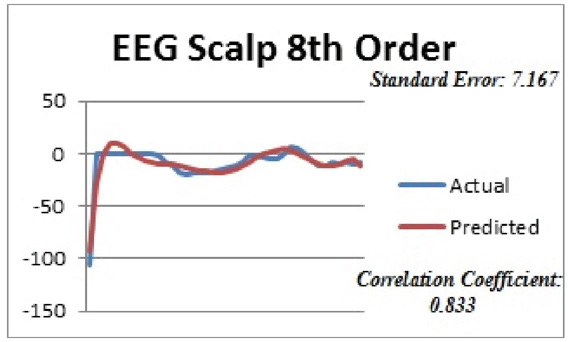
5.2 EEG Scalp Readings

The sample size of the data is 80, taken at 0.0039062 sec intervals. EEG signal measured is FZ-CZ, and readings are in uV.

5.2.1 4th Order



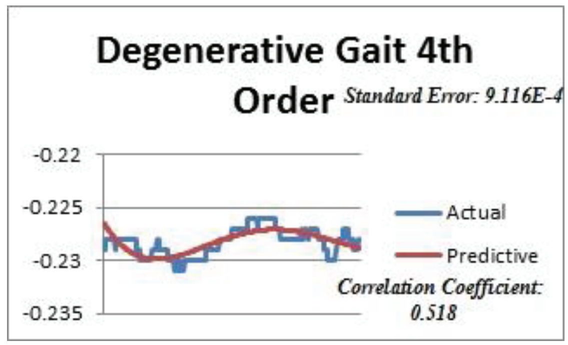
5.2.2 8th Order



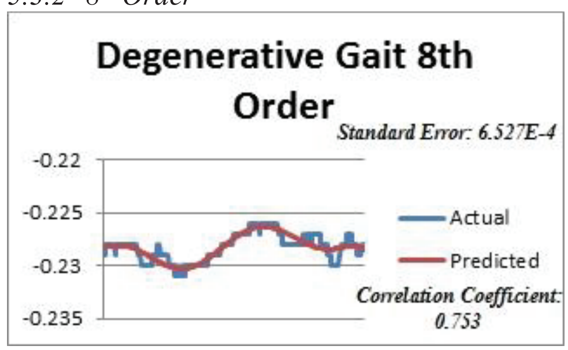
5.3 Degenerative Gait EEG Readings

The sample size is 300, and measures signals sent to left leg at 0.0033333 sec intervals. Readings are in mV.

5.3.1 4th Order



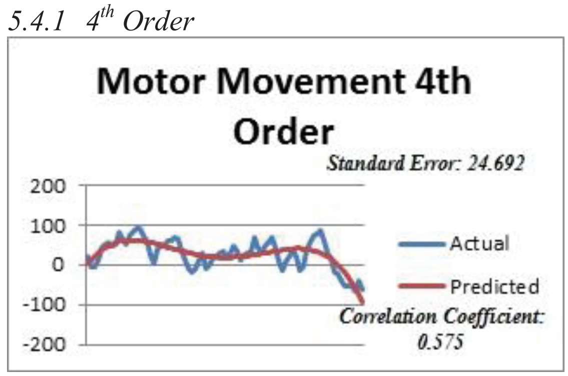
5.3.2 8th Order



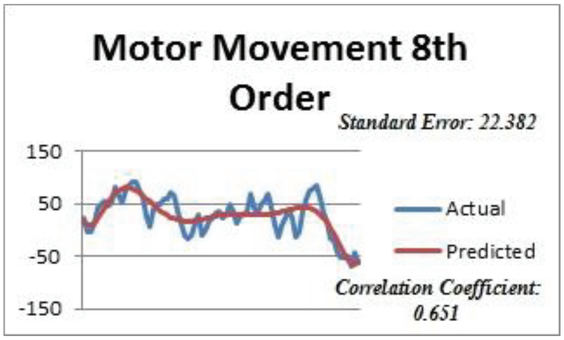
5.4 Motor Movement EEG Readings

Sample size is 80, with data taken in 0.00625 sec intervals. The measured signal is FCZ, and readings are values of uV.

5.4.1 4th Order



5.4.2 8th Order



6. INTERPRETATION & OBSERVATION

Our research attempts to model real world information using polynomials. In this case our model is the attempt to describe some real world phenomenon using mathematical concepts and languages. Essentially, if our polynomial reflects real data with a high degree of accuracy, then we can consider it an effective mathematical model. We used 4 different biological signals to order to assess the ability of a polynomial function to model real world data. The reason for choosing motor movement signals, degenerative gait EEG readings, EEG scalp readings, and blood pressure measurements was in order to measure a variety of biological functions. For instance, blood pressure measure is the force with which the heart pushes blood through the body. This is very different from the EEG electric signals that can be detected through a patients scalp. Motor movement and degenerative gait measurements are different from the scalp readings in that they relate to real muscle movement in a person's body. Therefore the readings have a simple cause and effect relationship (if the signal looks a certain way then a person's arm will flex). Lastly, motor movement readings are taken from healthy individuals and represent the healthy working movement and could possibly be compared to patients with degenerative disorders. By having a variety of measurement we can decide where polynomials are more effective at predicting data. Both the 4^{th} and 8^{th} order polynomials we choose are for two reasons. First they both give a good indication of how well real data can be represented with polynomials. Second it how the relative increases in accuracy the increasing orders can provide. Often increasing the order of the polynomial provides decreasing benefits as they become larger, and at a certain point, the increasing complexity of the calculation negates the small increase in accuracy. That having been said, 4^{th} and 8^{th} order polynomials were chosen because they provided good contrast and do a good job of demonstrating the accuracy of our research.

In looking at our results, in every case 8^{th} order polynomials are more accurate. This is most obvious in blood pressure measurements and scalp EEG measurements where the accuracy saw an increase of around 0.5. These two are also the most accurate as measured by their correlation coefficient, though the accuracy of the degenerative gait predictions is still fairly high. The reason for the differences is likely to be due to the smoother transitions for scalp EEG signals and blood pressure measurements. As we said earlier, polynomials work best where there is a high correlation between measurements taken at similar times. The more erratic seeming the signals are, the lower the less polynomial can accurately model the results.

CONCLUSION

In this paper, we propose a prediction based model for the emerging Wireless Body Area Network. This model basically deals with four different types of data which are very much relevant in recent medical field. Our regression polynomial model was applied to each and every data function and it helped to determine a predictive value of those functions using both the 4th order and the 8th degree polynomial regression function. This leads to significant saving of energy and time. Also as the numbers of nodes connected to a cluster head (CH) increases the amount of data the CH needs to pass on, remains constant as a factor of the order of the polynomial. Another benefit is that data outside the scope of the measured data can be predicted; assuming the polynomial accurately models the system.

FUTURE WORK

In the future work, we are trying to work with more biological functions using our model. With a larger variety of functions, it may be clearer where polynomials could be used to the greatest effect. Also we are trying to increase the number of samples so that the predictive values should be more prominent and clearly understandable for the users. Also, some of the signals (such as the motor movement signals) are quite erratic at a small scale. By increasing the scale we may find that the trends that were difficult to predict and seemingly erratic will become more obvious and predictable. We are trying to increase the accuracy of the polynomials so that the output would become more accurate and provide a robust model for prediction. The obvious way of doing this is to increase the order of the polynomial, though that would come at a cost of greater computing time and most data that needs to be sent, which is the opposite of our purpose. It would also be useful to compare this to popular methods of data aggregation to see where it is more or less effective.

REFERENCES

[1] Chakraborty, S., Ghosh, S. K., Jamthe, A., & Agrawal, D. P. (2013). Detecting mobility for monitoring patients with Parkinson's disease at home using RSSI in a Wireless Sensor Network. *Procedia Computer Science*, *19*, 956-961.

[2] Jamthe, A., Chakraborty, S., Ghosh, S. K., & Agrawal, D. P. (2013, May). An Implementation of Wireless Sensor Network in Monitoring of Parkinson's Patients Using Received Signal Strength Indicator. In *Distributed Computing in Sensor Systems (DCOSS), 2013 IEEE International Conference on* (pp. 442-447). IEEE.

[3] Torsha Banerjee, Kaushik R. Chowdhury, Dharma P. Agrawal: Distributed data aggregation in sensor networks by regression based compression. MASS 2005

[4] Torsha Banerjee, Kaushik R. Chowdhury, Dharma P. Agrawal: Using polynomial regression for data representation in wireless sensor networks. Int. J. Communication Systems 20(7): 829-856 (2007)

[5] Torsha Banerjee, Bin Xie, Jung Hyun Jun, Dharma P. Agrawal: Increasing lifetime of wireless sensor networks using controllable mobile cluster heads. Wireless Communications and Mobile Computing 10(3): 313-336

[6] Physionet: Available at http://www.physionet.org/cgi-bin/atm/ATM