CHAPTER 5

DEVELOPMENT OF ALGORITHM FOR DENOISING OF CGM SENSOR DATA

5.1 INTRODUCTION

Of late, smoothening of CGM sensor data has gained importance to facilitate its use in further signal processing applications like predictive alert generation for hypo/hyperglycemia and for optimal decision making for insulin dosage in closed loop control. The CGM devices assist the diabetic clinicians to analyze the fluctuations and their varying trend in BG levels. Evaluation of accuracy in CGM systems is very much complex, because of the indirect assessment of BG through ISF. This is associated with calibration using SMBG and the interdependence of data points that reflect the underlying process in time. Apart from the physiological time lag, improper calibration, random noise and errors due to sensor physics and chemistry also affect the accuracy of CGM data. This deteriorates the performance of CGM signals in hypo/hyperglycemic alert generation and suboptimal decision of control input to artificial pancreas. This chapter gives the details about HFT for denoising of CGM sensor data which improves the quality of input signal for the prediction approaches.

5.2 HYBRID FILTERING TECHNIQUE

The task of denoising is accomplished through the HFT which consists of a FFNN trained with an Extended Kalman Filter (EKF) algorithm. The EKF has been designed with the SS and measurement models obtained
through the dynamics of BG, IG and sensor gain deviation. The filtering capability of HFT has been compared with MA and Kalman Filter. The so obtained denoised signal is used as input for the prediction models for predicting the near future glucose concentrations in three different approaches.

In the proposed work of HFT, in the EKF algorithm, the minimum variance estimate of the state vector of a dynamic discrete time BG estimation process was governed by a nonlinear stochastic difference equation given in Equation (5.1)

\[ X(k+1) = f(X(k), w(k)) \]  

and with a measurement vector given in Equation (5.2),

\[ Y(k) = h(X(k), v(k)) \]

Where, ‘f ’ and ‘h ’ were nonlinear vector functions. \( W_k \) and \( V_k \) were the vectors of process and measurement noises respectively. \( \omega(k) \) and \( \nu(k) \) were assumed to be zero mean white noise processes with covariance matrices \( Q_k \) and \( R_k \) respectively. ‘f ’ was the nonlinear function, which related the state at previous time step ‘\( k - 1 \)’ to the state at time step ‘\( k \)’. ‘h ’ was the nonlinear function, which related the state and measurement vector. Estimation of state vector \( X \) has been done similar to a linear case. The time update equations were given in Equations (5.3) to (5.7) as follows.

Estimate of state vector:

\[ \hat{X} \left( \frac{k}{k} \right) = f \left( \hat{X} \left( \frac{k}{k} \right), 0 \right) \]  

(5.3)

Estimate of Covariance matrix at k’th sampling time:

\[ P \left( \frac{k+1}{k} \right) = A_k P \left( \frac{k}{k} \right) A_k^T + W_k Q_k W_k^T \]  

(5.4)

The measurement update equations were

\[ K_k = P \left( \frac{k+1}{k} \right) H_k^T \left( H_k P \left( \frac{k+1}{k} \right) H_k^T + V_k R_k V_k^T \right)^{-1} \]  

(5.5)

\[ \hat{X} \left( \frac{k+1}{k+1} \right) = \hat{X} \left( \frac{k}{k} \right) + K_k (Y(k + 1) - (\hat{X} \left( \frac{k+1}{k} \right), 0) \]  

(5.6)
\[ P\left(\frac{k+1}{k+1}\right) = (I - K_kH_k)P\left(\frac{k+1}{k}\right), 0 \]  
(5.7)

Where \( A_k \) and \( H_k \) were the Jacobian matrices of the partial derivatives of ‘f’ and ‘h’ with respect to \( X \), whereas \( W_k \) and \( V_k \) were the Jacobian matrices of partial derivatives of ‘f’ and ‘h’ with respect to \( w_k \) and \( v_k \). \( K_k \) is the Kalman gain matrix at the \( k \)th sample and \( I \) is the identity matrix with dimension as that of \( X \). The State transition matrix \( A_k \) obtained from the process model and the vector \( H_k \) from the measurement model were applied to the neural network as initial weights.

Development of mathematical models and their parameter estimation are essential in studying the dynamic behavior of biological systems. A natural way to model dynamic biological systems is to employ nonlinear SS equations. The SS representation of BG dynamics could be explained as follows. Any type of filter would try to obtain an optimal estimate of the desired quantities (the system’s state), from the data provided by a noisy environment. The concept of optimality corresponds to the minimization of the state estimation error in any respect. A SS model is any model, that includes an observation process \( Y_t \) and state process \( X_t \). The equations may be non-linear or Gaussian. It is usual in signal processing applications to use an Auto Regressive (AR) model to describe and predict various types of natural phenomena. The AR model attempts to predict the output of a system based on previous outputs. For an AR(2) process, the previous two terms \( X(t-1) \) and \( X(t) \) and the noise term \( w(t) \) contribute to the output \( X(t) \).

\[ X(t + 1) = \varphi_1 X(t) + \varphi_2 X(t - 1) + w(t) \]  
(5.8)

If the coefficients \( \varphi_1 \) and \( \varphi_2 \) are positive, then the output will resemble a low pass filter, with the decreased high frequency part of noise. If \( \varphi_1 \) is positive and \( \varphi_2 \) is negative, then the process favors changes in the sign, resulting in
oscillating output. Akaike Information Criteria (AIC) are a measure of the relative goodness of fit of a statistical model (Brockwell & Davis 2009).

\[
\text{AIC} = 2 \times K - 2 \times \ln (L) \tag{5.9}
\]

Where \( K \) is the number of parameters in the model and \( L \) is the maximized value of the likelihood function, for the estimated model. The preferred model is one with a minimum AIC value. The AIC penalizes for the addition of parameters, and thus, selects a model that fits well, but has a minimum number of parameters.

The process model to describe the physiological dynamics of IG given in Equation (5.10) was obtained from the work of Rebrin et al (1999). This model described the relation between IG and BG with a first order differential equation. The transport of glucose from blood to ISF has been represented as a diffusion model as given by,

\[
\frac{dIG(k)}{dk} = \left(\frac{1}{\tau}\right)IG(k) + \left(\frac{g}{\tau}\right)BG(k) \tag{5.10}
\]

Where \( dIG /dk \) was the rate of change of ISF glucose, \( \tau \) was the time constant, that defined the dynamic relationship between BG and IG and \( g \) was the static gain, equal to ‘1’ under steady state conditions.

\subsection*{5.3 PROPOSED ALGORITHM FOR DENOISING}

\textbf{Step1} : A Priori Estimates :

- \( X_k, X_k^\circ \)
- Error : \( e_k = X_k - X_k^\circ \)
- Process Noise Covariance : \( P_k = E( e_k e_k^T ) \tag{5.11} \)
- Measurement Noise Covariance : \( R_k \) (assigned from Noise model)

\textbf{Step2} : Obtain Initial Values : \( BG(k-1), BG(k), IG(k-1), IG(k), \hat{\alpha}(k-1), \hat{\alpha}(k) \)
Step3 : Estimate :
  o Parameters : $a_1, a_2, b_1, b_2$ through LS procedure.
  o Noise Variances : $\sigma_1^2, \sigma_2^2$ from the input signal.

Step4 : Compute :
  o SS model
  o Measurement model
  o Kalman Gain

From State equations
  • $A_{(6x6)} = df/dx$ --- Weights of input neurons to Hidden Layer. (5.12)
  • $H_{(1x6)} = dh/dx$ --- Weights of hidden neurons to Output Layer. (5.13)
  • $W_{(6x1)} = df/dw$ --- Bias to Hidden Layer neurons. (5.14)
  • $V_{(1x1)} = dh/dv$ --- Bias to Outer Layer neuron. (5.15)

Kalman Gain : $K_k = P_kH^T(HP_kH^T + R)^{-1}$ (5.16)

Let $\beta = n^*K_{ib}$ (5.17)

an adaptive parameter used to train the neural network according to
the variations in input signal.

Step5 : Train the NN with the computed SS and Measurement model through
EKF Algorithm.

Hidden Layer Activation Function : $f_1$ for
$$X_{k+1} = (A \times X_k) + W_k$$ (5.18)

Where $f_1(X, \beta) = x \times (x/(x - \beta))$ (5.19)

Where ‘$\beta$’ is the learning rate parameter derived from Kalman gain
and ‘$x$’ is the input to the hidden layer neuron which is equal to
the sum of products of inputs and their corresponding weights.

Outer Layer Activation Function : $F_0$ for
$$Z_k = (H \times X_k) + V_k$$ (5.20)

Where, $F_0(x, \beta) = x + (\beta \times x)$ (5.21)
Where ‘x’ is the input to output neuron which is equal to the sum of products of hidden layer outputs and the corresponding weight.

**Updation:**

\[ X_{k+1} = X_k + K_k (Z_k - (HI \times X_k)) \]  
(5.22)

\[ A_{k+1} = (X_{k+1})^T \times (X_{k+1}) \]  
(5.23)

Step6: Calculate RMSE between NN output and reference values.

Step7: Check for Optimization Criteria. If not met, repeat the steps 3-6

### 5.4 IMPLEMENTATION OF HFT

BG fluctuations are continuous processes in time $BG(t)$. Each point of that process is characterized by its location, speed and direction of change. Thus, at any point in time, $BG(t)$ is a vector. CGM sensors allow the monitoring of the process in short (e.g. 3 to 10 minutes) increments, producing a discrete time series that approximates $BG(t)$. The inputs with a sampling period of 5 minutes were considered, and hence, a set of 6 inputs could be gleaned from a 30 minute window, and given as input vector for the neural network. The moving window approach has been very much useful in observing the variations in SNR and to adapt the filter parameters accordingly. This window size has been conformed/optimized with false nearest neighbor method (Rhodes et al 1997) and the input vector size also coincided with state variable size.

Initially 2 BG sample values were taken as reference, with the help of One Touch Ultra glucometer, which measured the glucose concentration from the capillary blood obtained with a finger prick. Then, these values were applied to an AR model of order 2 to represent the dynamics of BG.

\[ BG (k + 1) = a1 \times BG (k) + a2 \times BG (k - 1) + w1 (k) \]  
(5.24)
Where, $BG(k)$ was the BG value at `k`th time instant, $wI(k)$ was the additive measurement error with variance $\sigma_1^2$. In the preprocessing stage of this research work, the variance $\sigma_1^2$ was found to be varying in the range of 2.5 to 7.6 mg$^2$/dL$^2$ and hence, an average value of 5.02 mg$^2$/dL$^2$ was used for further processing. Initially, the model order was taken empirically, and later, confirmed with AIC. From Equation 5.10, the ISF glucose at time step $k+1$ was,

$$lG (k + 1) = \left(1 - \frac{1}{\tau}\right) lG (k) + \left(\frac{1}{\tau}\right) BG$$

(5.25)

Where $lG (k)$ was the ISF glucose at time `k`, $BG(k)$ was the BG value at time `k`, ‘$g$’ was the gain and ‘$\tau$’ was the lag for the diffusion of glucose from ISF to blood.

The CGM value differed from the ISF glucose value with time varying sensor gain deviation ‘$\dot{a}$’. In the current research work, it has been observed from the results that, an AR(2) model itself is sufficient to describe the losses in performance of CGM sensor.

$$\dot{a}(k + 1) = c1 \ast \dot{a}(k) + c2 \ast \dot{a}(k - 1) + w2 (k)$$

(5.26)

where, $w2 (k)$ was a zero mean additive noise with variance $\sigma_2^2$. The coefficients used in Equations (5.24) and (5.26) were obtained with Ordinary Least Squares (OLS) method and verified with Yule Walker Equations.

To obtain a SS dynamic model for the estimation of glucose concentration in blood plasma, let $x1 = BG(k)$, $x2 = BG(k - 1)$, $x3 = lG(k)$, $x4 = lG(k - 1)$, $x5 = \dot{a}(k)$ and $x6 = \dot{a}(k - 1)$. Then the state equations for the estimation process could be written as follows.

$$x1(k + 1) = a1 \ast x1(k) + a2 \ast x2(k) + w1(k)$$

(5.27)

$$x2(k + 1) = x1(k)$$

(5.28)
The measurement model is given by,

\[ Y(k) = x3(k) * x5(k) + v(k) \]  \hfill (5.33)

Where, \( Y(k) \) was the CGM signal obtained with \( IG(k) \), with multiplicative sensor gain deviation \( \hat{\alpha}(k) \) and \( v(k) \) was the zero mean noise distribution with variance \( \sigma^2 \). The state matrix ‘\( A \)’ has been framed from Equations (5.27) to (5.32) and measurement vector ‘\( H \)’ from Equation (5.33). The extracted matrices were given below in Equations (5.34) and (5.35).

\[ A = \begin{bmatrix} 
\alpha_1 & a2 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 \\
1/ \tau & 0 & 1 - (1/ \tau) & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & c1 & c2 \\
0 & 0 & 0 & 0 & 1 & 0 
\end{bmatrix} \]  \hfill (5.34)

\[ H = \begin{bmatrix} 0 & 0 & 1 & 0 & 1 & 0 \end{bmatrix} \]  \hfill (5.35)

These state and measurement models were used for the initialization and training of Neural Network, that has been employed for denoising of errors in CGM sensor data. The neural network could model the input-output behavior of the glucose metabolism and with the assistance of EKF; it is applied for denoising of errors in glucose dynamics.
The proposed method was implemented and tested with all the data sets. The CGM values, thus obtained were mixed with the WGN of variance \( \sigma^2 = 5 \) to 52 mg\(^2\)/dL\(^2\). Though \( \tau \) varied with one user to another, with sensor location and with time, here it is kept to be fixed i.e., 6 minutes on the basis of literature. However, this uncertainty was compensated in real time with EKF applied to NN.

The process and measurement noise variances were initialized with \( q = 0.1 \) and \( r = 0.1 \) mg\(^2\)/dL\(^2\) respectively. The data sets were taken in such a way that each set corresponded to a 24 hour day with 5 minute sampling periods. Hence, each data set consisted of 288 IG values. Among these 288 values, 50\% were assigned for training, 25\% for validation and remaining 25\% for testing.

The architecture of NN used for HFT is shown in Figure 5.1. It was a multilayer feed forward back propagation neural network with an input, hidden and an output layer. In the feed forward stage, each input unit \((X_i)\) receives an input signal and transmits this signal with a weightage to each of the hidden units. Each hidden unit summarizes the inputs and its bias \((H_j)\), then calculates the activation function \((f_i)\) and sends its signal \(Z_i\) to each output unit. The output unit calculates the activation function \((F_o)\) with bias \((V)\) to form the response of the net for the given input pattern. These weights were decided by the EKF algorithm. During the back propagation of errors, each output unit compares its computed activation \(Y_k\) with its target value \(T_k\) to determine the error. Based on the RMSE, the factor \(\delta_k\) \((k = 1,2,\ldots,m)\) is computed and used to distribute the error at the output unit \(Y_k\) back to all units in the hidden layer. Similarly, the factor \(\delta_j\) is computed for each hidden unit \(Z_j\). Since the EKF provides minimum variance estimates in system applications where the dynamic process and measurement models contain non linear relationships, the non linearity in the glucose time series could very well be estimated and corrected by EKF.
As said earlier, the inputs to the neural network were applied as vectors of size $6 \times 1$. The state vector was of size $6 \times 6$. The noisy interstitial glucose values from the sensor were given to the network input layer. At hidden layer neural nodes, the weighted inputs were applied to a customized activation function, $f_1$, used to introduce nonlinearity in the model. The nodes at the output layer were applied with linear activation function, $F_o$. The output of the network depended upon the weights $A_{i,j}$ of the neurons, determined by the EKF state estimation and correction approach. ‘n’ represented iteration and the suffices $i,j$ denoted the traversal of input from $i$’th node in the first layer to $j$’th node in the next layer. $i = 1, 2,...N_i$ and $j = l, 2,...N_j$. In each iteration the state estimation and correction were carried out with the updated variance and Kalman gain values. The customized activation functions $f_1$ and $f_0$ respectively in the hidden and output layer, has made the network to adapt the parameters to meet out the performance metric of minimum RMSE to be less than 5 mg/dL. This adaptability accounted for the inter person and intra person variabilities of the signal to noise ratio. An important feature of the EKF is that the Jacobian in the

**Figure 5.1 Structure of neural network in HFT**
equation for the Kalman gain serves to correctly propagate or “magnify” only the relevant component of the measurement information. Since the HFT comprised of only one hidden layer, the computational complexity was less in the designed network. This NN trained with EKF confers an excellent convergence performance, compared to traditional GD optimization technique.

One of the representative glucose profile added with WGN is shown in Figure 5.2. The noisy CGM data were applied to ANN, trained with EKF algorithm. The filtering process was carried out in three phases, viz. training, validation and testing phase. The data were split in the ratio of 50%: 25%: 25% respectively for the three phases.

![Noisy CGM Data Graph](image)

**Figure 5.2 CGM sensor profile with WGN of variance 16 mg²/dL²**

The data were assigned according to dividerand function, which is default in the training of neural network. During the training phase, the weights and bias of neural nodes were estimated by EKF algorithm. The network was trained to meet the performance criteria that, RMSE between the actual value and filtered value to be less than 5 mg/dL. Updating of weights
and biases enabled the network to capture the nonlinearities in the CGM time series.

5.5 PERFORMANCE EVALUATION OF HFT

To analyze the denoising performance of the proposed technique, the HFT was evaluated with CGM signals of normal, hypo and hyperglycemic ranges. WGN, with variance ranging from 2 to 52 mg²/dL² has been introduced into the CGM signal. Different scenarios were generated with varied SNR in CGM sensor data, so as to produce interpatient variability and intraindividual variability. The interpatient variability is reflected through CGM signals of different SNR values. And intra individual variability is obtained by adding different noise variances at different time periods in a single CGM signal. The performance of HFT was compared with the MA and Kalman filter approaches. The denoising of CGM sensor data with online tuning of KF method used an initial tuning period of 6 hours, during which, the process and measurement variance parameters were estimated. This work has been extended to account for interindividual and intraindividual variability of SNR. i.e., the noise variance parameters were estimated, whenever there was a variation in SNR value. The qualitative and quantitative analysis of HFT is given below.

The HFT provided perfect smoothing by the intelligence of sensing the fluctuations in variance of noise distributions, and other non physiological variations of BG dynamics. The results clearly proved that, the HFT mechanism was able to capture and remove the combined noise effects. This implied that, the artificial neural network in HFT was trained well with physiological variations of each individual, and it tracked the glucose profile perfectly, neglecting the various noise effects in CGM time series. The
denoising effect of HFT in a representative noisy glucose profile is illustrated in Figure 5.3. The data sets were represented as moving window of size 6 and sent as inputs for the neural network. For each input vector, the EKF algorithm predicted the next value and estimated the parameters with the help of state transition and measurement vector.

Figure 5.3 Performance of HFT in hypoglycemic range

The Kalman gain and the activation functions of the hidden and output layers stimulated the algorithm to make accurate estimations. RMSE between the filtered signal and the true glucose signal was calculated as the performance metric. From the trials conducted with 1400 data sets, it has been observed that the HFT method produced minimum RMSE, compared with KF in almost all the trials.
Figure 5.4 Performance of HFT in normoglycemia

Figure 5.5 Performance of HFT in hyperglycemic range
Exemplary results have been reported in Figures 5.3, 5.4 and 5.5 respectively, for the three ranges. It could be very well understood that, the HFT was able to catch up the dynamics of the incoming signal in all the three ranges. To analyze the performance of HFT with variable SNR, CGM profiles of different variance values were created. It was observed that the HFT coped well with variable SNR in CGM time series. The process and measurement covariance estimates, Kalman Gain and the activation functions of NN were updated and accommodated well with EKF algorithm in denoising the errors of variable noise parameters. From the experiments made, it has been observed that the HFT could provide optimal filtering with inter individual variability of SNR. Glucose profiles of different noise variances were created in a single time series at different time durations and applied as inputs to HFT, to check with the intra individual variability of SNR. Results have confirmed that, various noise models could be eliminated by HFT, with the hybrid of FFBP neural network trained with non linear EKF algorithm. The nonlinear physiological variations could be tracked by HFT with its adaptive estimation of parameters. A Comparison of HFT with KF is presented in Figure 5.6. The proposed HFT has also been tested with three days data set as shown in Figure 5.7. The HFT is able to denoise the errors even in three days’ data set with an increase in error percentage of less than 5 in terms of RMSE.
RMSE in mg/dL was calculated for MA, Kalman and Hybrid Filters. The three methods provided the results with Mean ± SD as 9.21 ± 4.3, 7.99 ± 3.6 and 5.38 ± 4.5 mg/dL respectively. The delays introduced in these methods were of an average 7.4, 5.2 and 3.95 minutes respectively. The amount of time needed for the estimation of neural weights with respect to the varying noise variances, could be the reason for the delay in the proposed HFT. SRG was high with an average value of 0.9, with the intelligence of neural network. The experiments were conducted for different scenarios as explained above, and the average results of each of the data sets have been compared with MA and KF in terms of RMSE, Time Delay and SRG, as listed respectively in Tables 5.1, 5.2 and 5.3.
Figure 5.7 Performance of HFT in Three days Data set

Table 5.1 Comparison of HFT with MA and KF in terms of RMSE

<table>
<thead>
<tr>
<th>Data Set</th>
<th>RMSE (Mean± SD) mg/dL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MA Filter</td>
</tr>
<tr>
<td>Data Set 1</td>
<td>8.8±2.5</td>
</tr>
<tr>
<td>Data Set 2</td>
<td>8.34±4.5</td>
</tr>
<tr>
<td>Data Set 3</td>
<td>10.5±3.5</td>
</tr>
</tbody>
</table>
Table 5.2 Comparison of HFT with MA and KF in terms of Time delay

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Time Delay (Mean± SD) Minutes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MA Filter</td>
</tr>
<tr>
<td>Data Set 1</td>
<td>7.0±2.33</td>
</tr>
<tr>
<td>Data Set 2</td>
<td>6.81±2.24</td>
</tr>
<tr>
<td>Data Set 3</td>
<td>8.38±4.1</td>
</tr>
</tbody>
</table>

Table 5.3 Comparison of HFT with MA and KF in terms of SRG

<table>
<thead>
<tr>
<th>Data Set</th>
<th>SRG (Mean± SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MA Filter</td>
</tr>
<tr>
<td>Data Set 1</td>
<td>0.69±0.09</td>
</tr>
<tr>
<td>Data Set 2</td>
<td>0.71±0.12</td>
</tr>
<tr>
<td>Data Set 3</td>
<td>0.68±0.14</td>
</tr>
</tbody>
</table>

Table 5.4 Overall analysis of denoising techniques

<table>
<thead>
<tr>
<th>AVERAGE VALUES</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE in mg/dL</td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>MA KF HFT</td>
</tr>
<tr>
<td>9.21 7.99 5.38</td>
</tr>
</tbody>
</table>
To analyze the performance of HFT, the overall average values of the performance metrics were compared for MA, Kalman Filter and HFT as given in Table 5.4. Performance comparison of the denoising techniques has been depicted in Figure 5.8. Considering the RMSE metric, HFT provided an improvement of 41.5% and 32.6%, when compared respectively with MA and Kalman filters. In the case of Time lag, an enhancement of 46.7 % and 25 % were obtained with HFT, compared to MA and KF respectively. The SRG in HFT offered an improvement of 29.8% and 9.3% respectively, with MA and Kalman filters.

5.6 SUMMARY

In this chapter, the SS model has been developed with the modeling of BG and sensor gain deviation with an AR model of order 2. The SS model and measurement model so obtained with EKF algorithm were used to train a
FFNN to track the non linear dynamics of CGM sensor data. And the denoising capacity of this HFT has been compared with the MA and Kalman Filter in terms of RMSE, Time lag and SRG. The denoised CGM sensor data would be used as input for the prediction approaches. The next chapter elaborates the CGM time series prediction with ARIMA model.