Using Mixtures of Experts on the GlucoWatch (R) Biographer

Jonathan J. Oliver jono@ultimode.com
Steve R. Waterhouse stevew@ultimode.com
Ultimode Inc., 2201 Broadway, Rm 215, http://www.ultimode.com
Oakland, CA 94612

Ronald T. Kurnik Ronald_Kurnik@cygn.com
Russell O. Potts Russ_Potts@cygn.com
Cygnus, Inc., 400 Penobscot Dr., http://www.cygn.com
Redwood City, CA 94063

October 12, 1999

1 Introduction and Scope

Glucose monitoring helps diabetic patients manage their blood glucose levels. Diabetics who carefully control their blood glucose levels can significantly reduce the number of long-term complications in diabetes [3].

Presently, diabetics estimate their blood glucose level by obtaining a small amount of blood with a finger stick, followed by external analysis of this blood sample for glucose content; the external analysis of the blood sample is performed by devices such as the HemoCue (R) clinical analyzer (HemoCue, AB, Sweden)[1].

The finger stick method is accurate, but it is also painful and inconvenient, particularly when the patient must determine blood glucose levels multiple times per day. In addition, the irregular timing of the finger stick method may result in an increased incidence of hypoglycemia[2] [3]. Continuous noninvasive measurement[3] of glucose is expected to reduce the long-term complications of diabetes, while minimizing hypoglycemia and inconvenience.

[1] See http://www.hemocue.se

[2] A hyperglycemic event is the blood glucose level going too high; a hypoglycemic event is the blood glucose level going too low.

[3] A noninvasive measurement avoids the need to extract blood.

This paper focuses on the statistical methods used to make estimates of blood glucose levels from noninvasive measurements. The statistical method described in this paper has been implemented on the GlucoWatch (R) biographer[4].

2 A Description of the GlucoWatch (R) Biographer

The biographer obtains noninvasive measurements by applying a low-level electrical current through the skin between an anode and cathode. Due to the applied potential, sodium and chloride ions (from beneath the skin) migrate towards the cathode and anode, respectively [6,7]. The glucose molecules (which are not charged) are carried along with the ions by convective transport [8]. Over the range of current densities used by the biographer ($0 - 0.5\ mA/cm^2$), glucose extraction is approximately linear with current density and duration of current [12, 4].

In typical use, a diabetic wears the biographer for 15 hours. After 3 hours, the finger stick method is used to get an estimate of the blood glucose level. This blood glucose level is entered into the biographer to calibrate later predictions. For the next 12 hours, the biographer records at 20 minute intervals two measurements of the the integrated current measured in nanoCoulombs (nC). Software incorporated into the device uses (a) the measurements made and (b) the calibration blood glucose level to inform the diabetic of their blood glucose level.

3 The Problem

We viewed the problem posed by the biographer as a multivariate regression problem where the input variables were the sensor readings and the target variable was $\hat{BG}$, the blood glucose level.

3.1 Approaches to Multivariate Regression

There are many approaches to multivariate regression including:

- Linear regression;
- K-nearest neighbour (KNN) [5, 10];
- Neural networks [13, 11];
- Multivariate adaptive regression splines (MARS) [6]; and
- Mixtures of Experts (MOE) [14, 9].

[4]See http://www.cygn.com/glucowatch.html

We wanted to establish which of these methods (if any) were suitable for making predictions. Linear regression, the KNN method and neural networks are well known and understood methods. We will briefly describe MARS and MOE methods.

3.1.1 MARS

Multivariate adaptive regression splines (MARS) developed by Friedman [6] builds a model which has the equational form of a sum of products of univariate splines. The input space is partitioned into overlapping regions; in each region a univariate spline is chosen to fit the training data. The model is built in two phases; a growing phase where the input space is partitioned and the splines are selected and a pruning phase which uses cross validation to remove regions and splines which do not generalise well.

3.1.2 Mixtures of Experts (MOE)

The Mixtures of Experts (MOE) [14, 9] builds a model which is a mixture of linear models. If we have M independent variables $x_1,\ x_2,\ \ldots x_M$ and dependent variable y a MOE model with N experts make predictions using an equation of the form:

$$\hat{y} \ = \ w_1 y_1 \ + \ w_2 y_2 \ + \ \ldots \ + \ w_N y_N$$

where each y_i is the result of a linear function:

$$y_i \ = \ a_{1i} x_1 \ + \ a_{2i} x_2 \ + \ \ldots \ + \ a_{Mi} x_M$$

and the w_i (the weights we associate with each expert) are calculated using a logistic method:

$$\begin{aligned} \text{Let } d_i \ &= \ b_{1i} x_1 \ + \ b_{2i} x_2 \ + \ \ldots \ + \ b_{Mi} x_M \\ w_i \ &= \ \frac{e^{d_i}}{\sum_{j=1}^{N} e^{d_j}} \end{aligned}$$

Training a MOE model requires that we estimate the N the number of experts, the a_{ij} and the b_{ij}.

3.2 The Input Variables

The software has 3 inputs variables:

- t — the time of the reading;
- $CurrentA_t$ — site A's integrated current (nC) at time t; and
- $CurrentB_t$ — site B's integrated current (nC) at time t.

At the time of calibration ($t = t_{Cal}$), the biographer records the these inputs (t_{Cal}, $CurrentA_{Cal}$ and $CurrentB_{Cal}$) and BG_{Cal} — the blood glucose level at calibration.

3.3 Derived Variables

We found it useful to define the following variables:

- the average of the two currents at time t

$$Average\_Current_t \;=\; \frac{CurrentA_t + CurrentB_t}{2}$$

- the time since the biographer was calibrated

$$t' \;=\; t - t_{Cal}$$

- $Signal_t$ where the signal at time t is defined as:

$$Signal_t = \frac{BG_{Cal} * (Average\_Current_t + \Omega)}{Average\_Current_{Cal} + \Omega} \tag{1}$$

  where Ω is a constant.

3.4 Evaluating a Blood Glucose Prediction Method

It is essential that a diabetic wearing the biographer is confident that the predictions are accurate. To determine whether a predictive method was accurate, we obtained a set of measurements from the biographer. In addition, we obtained reference blood glucose measurements using the HemoCue (R) analyzer.

To evaluate a predictive method we

1. split the data obtained into training and test sets;
2. trained the predictive method using the training set; and
3. evaluated the criteria (listed below) on the test set.

Let us consider a test set of N predictions $\hat{BG}_i$ with corresponding reference values BG_i. There were criteria which quantify how accurately the $\hat{BG}_i$ were:

- Mean absolute relative error (MARE);
- The percentage of predictions which are considered clinically accurate/acceptable and clinically significant errors;
- The percentage of predictions which are accuracte when the blood glucose is in the region which is considered clinically important; and
- The goodness of fit when we fit a linear regression to $\hat{BG}_i$ and BG_i.

In this paper, we considered a range of statistical criteria. First, no single criteria appeared to capture the quality of a prediction method, while the 4 criteria together gave a strong impression of the quality of a prediction method. Secondly, authorities (including statistical consultant experienced in medical devices and medical practitioners) cannot agree as to the relative importance of the criteria.

We now define each criteria in turn.

3.4.1 Mean Absolute Relative Error

The MARE is defined as

$$MARE = \sum_{i=1}^{N} \frac{|\hat{BG}_i - BG_i|}{BG_i}$$

We prefer to use the absolute relative error to the absolute error since the clinical importance of an error is greater when the blood glucose level is low. For example, a prediction error of 5 mg/dl when the true blood glucose level is 50 mg/dl is approximately equivalent to a prediction error of 20 mg/dl when the true blood glucose level is 200 mg/dl.

3.4.2 The Percentage of Clinically Accurate / Acceptable Predictions

Clarke [2] defines five types of predictions: A, B, C, D and E (as shown in Figure 1). Predictions which fall into region A are considered clinically accurate; predictions in region B are considered clinically acceptable; and predictions in region C-E are considered clinically significant errors.

3.4.3 The Accuracy of Clinically Important Predictions

Since it is more important to make accurate predictions when the blood glucose is low, we define a measure which captures the ability of a prediction method to make accurate predictions for low measurements. For this measure we define the clinically important region[5] as measurements where the reference is less than 100 and we define a prediction as being accurate when the prediction is within 20 mg/dl of the reference[6]:

$$\%Pred \pm 20 = \frac{Count(|\hat{BG}_i - BG_i| \leq 20 \quad AND \quad BG_i < 100)}{Count(BG_i < 100)} * 100$$

3.4.4 The Goodness of Fit when we Fit a Linear Regression to $\hat{BG}_i$ and BG_i

Two measures are the (a) slope and (b) intercept of a line which is generated by:

1. plotting the reference BG on the X-axis, the biographer predictions on the Y-axis; and
2. fitting a line by minimizing residual sum of squares in the Y direction.

We prefer biographer prediction methods which result in slopes close to 1, and intercepts close to 0.

[5] It is standard to use a 100 mg/dl level as the threshhold for the clinically important region.

[6] The 20 mg/dl tolerance level used here is a standard clinical number to use.

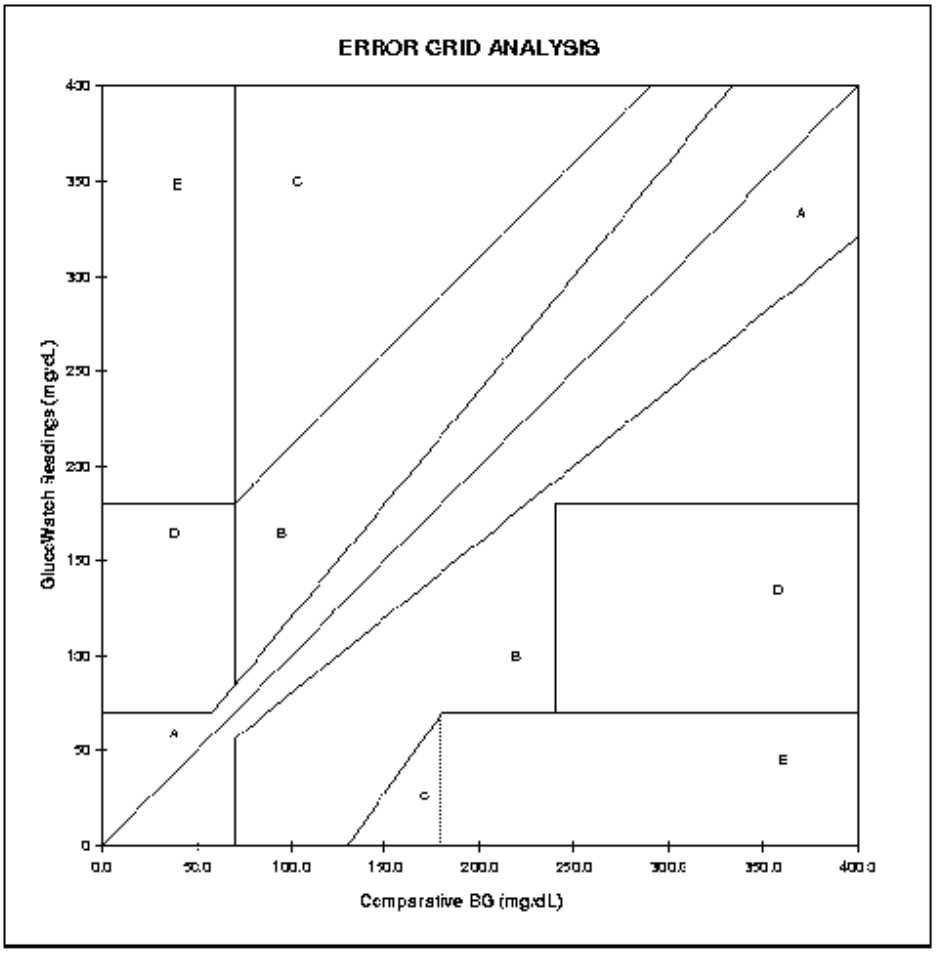

Figure 1: The Error Grid

However, the reference BG values are themselves a measurement made by the HemoCue (R) analyzer which has potential error. Since there are potential errors in both the HemoCue (R) analyzer readings and the biographer reading, one should use methods such as orthogonal regression which minimizes the distance to the line for each point [7, 1] and quote the slope and intercept of the orthogonal line. These measurement error models can be characterized by a parameter

δ - the ratio of the error variances associated with the two readings.

In this paper, we quote the slope and intercept using (a) standard least squares regression and (b) orthogonal regression with $\delta = 2$.

In addition, we use the coefficient of determination between $\hat{BG}_i$ and BG_i as a measure of the goodness of a prediction method.

4 Testing the Prediction Methods

Prediction Method	%MARE	%A	%B	%C	%D	%E
Neural Net 20 HU	18.58	64.3	30.1	1.1	4.4	0.0
Neural Net 40 HU	18.88	62.8	32.0	0.5	4.6	0.0
linear regression	22.57	52.9	41.7	0.8	4.6	0.1
MARS	14.92	74.3	22.1	0.3	3.2	0.0
KNN	23.14	53.6	38.7	0.0	7.7	0.0
MOE (3 experts)	14.43	74.0	22.8	0.6	2.7	0.0

Table 1: The results of the selected methods. The training set had 1417 points and the test set had 2888 points. The row labelled "Neural Net 20 HU" lists the results of a neural net with 20 hidden units and the row labelled "Neural Net 40 HU" lists the results of a neural net with 40 hidden units. The "% MARE" column lists the mean absolute relative error %. The columns %A – %E list the percentage of points which fell into each region according to the error grid in Figure 1.

Cygnus Inc., (Redwood City, CA) organized clinical trials where diabetic subjects wore the biographer over a 15 hour period. Every 20 minutes, the biographer recorded the time, integrated current and the blood glucose value measured by the HemoCue (R) analyzer (as described in Section 3.2). The set of measurements were split into training and test sets, we trained the statistical methods (listed in Section 3) and evaluated the criteria (from Section 3.4).

Tables 1 and 2 give the results when the selected prediction method were trained and tested on data collected in a development trial. The training set consisted of the measurements collected from 46 subjects wearing the biographer for up to 15 hours (a total of 1417 data points). The test set consisted of the measurements collected from 91 subjects wearing the biographer for up to 15 hours (a total of 2888 data points). In addition to the results listed, we generated results using MOEs with the numbers of experts ranging from 2 to 5; and neural nets with the number of hidden units varying from 20 to 100. We found that using the MOE with 3 experts gave highly accurate results; the neural net results with different numbers of hidden units were similar to the results in these tables.

4.1 Discussion of the Results

We summarize the results in Tables 1 and 2 as follows:

- The MARS method and Mixtures of Experts (MOE) appeared to be the most appropriate method for predicting the blood glucose level.

Prediction Method	Pred ±20	%Pred ±20	Slope	Inter-cept	Slope $\delta = 2$	Inter-cept $\delta = 2$	R2
Neural Net 20 HU	228	55.88	0.94	11.5	1.06	-7.4	0.72
Neural Net 40 HU	222	54.41	0.90	15.5	1.01	-1.9	0.71
Linear Regression	186	45.59	0.91	9.3	1.07	-14.8	0.66
MARS	281	68.87	0.88	18.8	0.94	8.9	0.80
KNN	201	49.26	0.54	65.3	0.61	54.4	0.52
MOE (3 experts)	312	76.47	0.96	2.2	1.04	-10.7	0.79

Table 2: Further results of the selected methods. The "Pred ± 20" column lists the number of points when the reference blood glucose was less than 100 mg/dl, and the prediction was within 20 mg/dl of the reference value. In the test set there were 408 points when the reference blood glucose was less than 100 mg/dl. The "% Pred ± 20" column lists the percentage of points when the reference blood glucose was less than 100 mg/dl, and the prediction was within 20 mg/dl of the reference value. The "Slope" and "Intercept" columns list the slope and intercept of a line fitting the predictions to the reference blood glucose using standard least squares regression; the "Slope $\delta = 2$" and "Intercept $\delta = 2$" columns list these values for the orthogonal regression. The "R2" column lists the coefficient of determination between the predictions and the reference values.

- The MOE method was superior to MARS for the prediction of blood glucose when the blood glucose was in the clinically important regions (Table 2).

4.2 Advantages of using MOEs on the Biographer

In addition to been an accurate method, the MOE prediction method has other advantages for implementation on the biographer:

- The MOE method can operate within the computational power and memory available on the biographer[7]. In contrast, the MARS method generates a model with many more parameters; and the KNN method requires that the biographer store the original training set.

- The predictions can be explained; Since, glucose extraction is approximately linear with current density and duration of current [12, 4], we expect the coefficients for $Signal_t$ (defined in Equation 1) should be close to 1. The coefficients in Equations 2 - 4 were in fact 1.0850, 1.0685 and 0.7449.

- MOEs scale well to larger training sets (from further clinical trials).

[7] Currently, the biographer can store approximately 50 real valued parameters.

5 Conclusion

We investigated a number of methods for predicting blood glucose levels on the biographer, including linear regression, neural networks, Multivariate Adaptive Regression Splines (MARS), Nearest Neighbour (KNN) and the Mixtures of Experts (MOE) approach. The preferred prediction method was the Mixtures of Experts (MOE) method with three experts. The MOE method (as defined in Appendix 1) has been implemented on the biographer. The biographer yields automatic measurements of glucose (up to 3/hr) over a 12 hour period with accuracy and precision comparable to existing, single point blood measuring devices.

Acknowledgments

We would like to thank Neil Ackerman, Rohan Baxter, Wray Buntine, Tim Dunn, Yalia Jayalakshmi, Matt Lesho, Margarita Lopatin, Janet Tamada and Charles Wei for their enthusiasm on this project and valuable discussions.

References

[1] G. Casella and R.L. Berger. *Statistical Inference.* Brooks/Cole, 1990.

[2] W.L. Clarke, D.C. Cox, L.A. Conder-Frederick, W. Carter, and S.I. Pohl. Evaluating clinical accuracy of systems for self monitoring of blood glucose. *Diabetes Care*, 10:622–628, 1987.

[3] The Diabetes Control and Complications Trial Research Group. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin dependent diabetes mellitus. *New England Journal of Medicine*, 329:977–1036, 1993.

[4] S.M. Dinh, C.W. Luo, et al. Upper and lower limits of human skin electrical resistance in iontophoresis. *AlChE Journal*, 39(12):2011–2018, 1993.

[5] R. O. Duda and P. E. Hart. *Pattern Classification and Scene Analysis.* John Wiley and Sons, New York, 1974.

[6] J.H. Friedman. Multivariate adaptive regression splines. *Annals of Statistics*, 19:1–141, 1991.

[7] W.A. Fuller. *Measurement Error Models.* Wiley, New York, 1987.

[8] P. Glikfeld, R.S. Hinz, et al. Noninvasive sampling of biological fluids by iontophoresis. *Pharm. Res.*, 6:988–990, 1989.

[9] M. Jordan and R. Jacobs. Hierarchical mixtures of experts and the EM algorithm. *Neural Computation*, 6:181–214, 1994.

[10] D.G. Lowe. Similarity metric learning for a varable–kernel classifier. *Neural Computation*, 7:72–85, 1995.

[11] David J.C. MacKay. A practical Bayesian framework for backpropagation networks. *Neural Computation*, 4:448–472, 1992.

[12] M.J. Pikal. The role of electrosmotic flow in transdermal iontophoresis. *Adv. Drug Del. Rev.*, 9:201–277, 1992.

[13] D.E. Rumelhart, J.L. McClelland, and the PDP Research Group. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume I Foundations.* MIT Press, Cambridge, MA, 1986.

[14] S.R. Waterhouse and D.J.C. MacKay. Bayesian methods for mixtures of experts. In D.S. Touretzky, M.C. Mozer, and M.E. Hasselmo, editors, *Advances in Neural Information Processing Systems*, volume 8, pages 351–357. MIT Press, 1996.

Appendix 1: The MOE Set of Equations

The MOE method implemented on the biographer requires the 4 variables defined in Section 3.3: t', $Average\_Current_t$, BG_{Cal} and $Signal_t$.

We use the following process to make predictions:

1. Calculate d_1, d_2 and d_3:

$$\begin{aligned} d_1 &= 0.1943\ t' + 0.00003854\ Average\_Current_t \\ &\quad - 0.00128\ Signal_t - 0.00411\ BG_{Cal} - 0.9702 \\ d_2 &= 0.0641\ t' - 0.00010713\ Average\_Current_t \\ &\quad + 0.00622\ Signal_t + 0.00726\ BG_{Cal} - 1.0287 \\ d_3 &= -0.2584\ t' + 0.00006859\ Average\_Current_t \\ &\quad - 0.00494\ Signal_t - 0.00315\ BG_{Cal} + 1.9989 \end{aligned}$$

2. Calcuate w_1, w_2 and w_3 (the weights for each expert):

$$w_1 = \frac{e^{d_1}}{e^{d_1} + e^{d_2} + e^{d_3}} \qquad w_2 = \frac{e^{d_2}}{e^{d_1} + e^{d_2} + e^{d_3}} \qquad w_3 = \frac{e^{d_3}}{e^{d_1} + e^{d_2} + e^{d_3}}$$

3. Calculate $\hat{BG}_1$, $\hat{BG}_2$ and $\hat{BG}_3$ (the predictions made by each expert):

$$\begin{aligned} \hat{BG}_1 &= 6.1296\ t' + 0.00343065\ Average\_Current_t \\ &\quad + 1.0850\ Signal_t - 0.1150\ BG_{Cal} - 19.3839 \end{aligned} \tag{2}$$

$$\begin{aligned} \hat{BG_2} &= 5.6570\ t' + 0.00471199\ Average\_Current_t \\ &\quad + 1.0685\ Signal_t + 0.0827\ BG_{Cal} - 44.6528 \end{aligned} \tag{3}$$

$$\begin{aligned} \hat{BG_3} &= 4.3137\ t' + 0.00184351\ Average\_Current_t \\ &\quad + 0.7449\ Signal_t - 0.0484\ BG_{Cal} + 10.5759 \end{aligned} \tag{4}$$

4. Make a prediction as the weighted average of the predictions made by each expert:

$$\hat{BG} = w_1 BG_1 + w_2 BG_2 + w_3 BG_3$$