31st Annual International Conference of the IEEE EMBS
Minneapolis, Minnesota, USA, September 2-6, 2009

Piecewise-Linear Trend Detection in Longitudinal Physiological Measurements

Stephen J. Redmond, *Member*, *IEEE*, Jim Basilakis, Yang Xie
Branko G. Celler, *Member*, *IEEE*, Nigel H. Lovell *Senior Member*, *IEEE*

Abstract— **Recently, telecare solutions have been demonstrated as an effective means of monitoring chronic disease at a distance. A clinician may be managing many tens or hundreds of remote patients, prompting the need for a decision support system (DSS) to provide a more automated approach to managing these vast amounts of data. While simple threshold-based alert techniques provide some utility in notifying clinicians of extreme out-of-range parameter values, more incipient changes in a subject's condition may be sooner recognized by identifying trends in the longitudinal parameter data. Here we describe an approach for obtaining a piecewise-linear fit, to longitudinal physiological trend data, comparable with a similar fitting performed by a human observer, using a graphical user interface. The technique has been applied to both simulated and real data, and a comparison performed against the human scoring for each. On simulated data, the method matches or betters the human performance in most cases; with the greatest improvement observed in more noisy data. Similarly, for real physiological data, the deviation from the human marking, as a fraction of total variability of the signal, is less than 0.35.**

I. INTRODUCTION

Recently, telecare solutions have been demonstrated as an effective means of monitoring chronic disease at a distance [1]. Diseases, such as chronic obstructive pulmonary disease (COPD) and congestive heart failure (CHF), can be observed through the unsupervised measurement of various physiological parameters; namely, electrocardiogram (ECG), blood pressure (BP), spirometry, pulse oximetry, weight and temperature; and increasingly a number of more invasive measurements, like blood coagulation times (INR).

The ability to obtain frequent updates of a patient's physiological condition provides an opportunity to identify health deterioration at an earlier time-point and preemptively address any issues before hospitalization is required. However, it is the high frequency of data measurement (approximately daily) which highlights a new challenge in the interpretation of these data.

A clinician may be managing many tens or hundreds of patients. With each patient generating several physiological parameters from each of the biosignal mentioned earlier, the clinician will soon become overloaded with data. This prompts the need for a decision support system (DSS) to provide a more automated approach to managing these vast amounts of data.

This work was supported by the Australian Research Council.

S. J. Redmond, Y. Xie and N. H. Lovell are with the Graduate School of Biomedical Engineering, University of New South Wales, NSW 2052, Sydney, Australia. (n.lovell@unsw.edu.au)

J. Basilakis and B. G. Celler are with the School of Electrical Engineering and Telecommunications, University of New South Wales.

While simple threshold-based alert techniques provide some utility in notifying clinicians of extreme out-of-range parameter values [2], more incipient changes in a subject's condition may be sooner recognized by identifying trends in the longitudinal parameter data. While many data analysis techniques exist, which may denoise the data and uncover the underlying data trend, the authors believe that the interpretation of the data by a clinician is made more accessible via a piecewise-linear fitting; in which case the data may be summarized by the rate of change and the total duration of the most recent trends.

This paper describes a technique for obtaining a piecewise linear fit to data, similar to that obtained from a human scorer marking the same data. The algorithm is tested on both simulated data and real physiological data obtained from a one year telehealth trial.

II. METHODS

A. Data Set

1) Simulated data: Simulated data were generated to test, in a controlled manner, the ability of the proposed piecewise-linear regression to track the true underlying signal. The simulated data attempts to mimic some commonly occurring artifacts anecdotally observed in longitudinal physiological measurement data, such as: varying degrees of inherent variability in the measurement values; extreme outliers, caused by poor measurement technique or erroneous parameter extraction (such as deriving a heart rate from an ECG); and extended data outages, which occur when the patient is admitted to hospital for several weeks, or when the patient simply stops performing recordings for an extended period.

To simulate the underlying signal, over a one year period, a continuous sequence of linear trends is constructed. The y-value of the starting point of the first trend is randomly chosen with uniform probability on the range $[-1, +1]$ (no units). The final y-value for the trend is dependent on whether the initial value, let us term it y_{start}, was greater than, or less than zero; if it less than zero the final value is randomly selected with uniform probability on the interval $[y_{start}, +1]$, otherwise it is similarly chosen on the interval $[-1, y_{start}]$. This selection process provides some variability in the simulated signal, while ensuring it is bounded to the $[-1, +1]$ interval. The duration of each trend segment is selected randomly with uniform probability on the interval $[1, 84]$ days. This generated sequence of concatenated linear trends, is sampled every minute over its duration; this signal

$$\begin{bmatrix} x_1 & (x_1-b_1)_+ & (x_1-b_2)_+ & \cdots & (x_1-b_K)_+ & 1 \\ x_2 & (x_2-b_1)_+ & (x_2-b_2)_+ & \cdots & (x_2-b_K)_+ & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ x_N & (x_N-b_1)_+ & (x_N-b_2)_+ & \cdots & (x_N-b_K)_+ & 1 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_K+1 \\ w_{K+2} \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} \quad (1)$$

$$\mathbf{X}\mathbf{w} = \mathbf{y} \quad (2)$$

is later decimated to obtain approximately one measurement value per day.

Two forms of noise are now added. Firstly, white Gaussian noise, with a standard deviation of $\sigma_N \in \{0.01, 0.05, 0.1, 0.5, 1.0\}$, is added to every sample. Next, white Gaussian noise, with a standard deviation of $10\sigma_N$ is added to a configurable fraction, $p_N \in \{0, 0.05, 0.1, 0.15, 0.2\}$, of randomly selected points. To simulate a sustained outage in measurement values, a configurable fraction, $p_o \in \{0, 0.05, 0.1, 0.15, 0.2\}$, of consecutive samples are removed. Finally, a random selection of data points are taken such that approximately one data point per day occurs; although the selection process does not exclude the possibility of several simulated measurements occurring in the same day, or similarly no measurements occurring for several days. Varying σ_N, p_N and p_o across the set of values shown for each results in 125 simulated data signals. Fig. 1 provides an illustrative example of one such signal, using the following choice of parameters: $\sigma_N = 0.1$, $p_N = 0.2$ and $p_o = 0.2$.

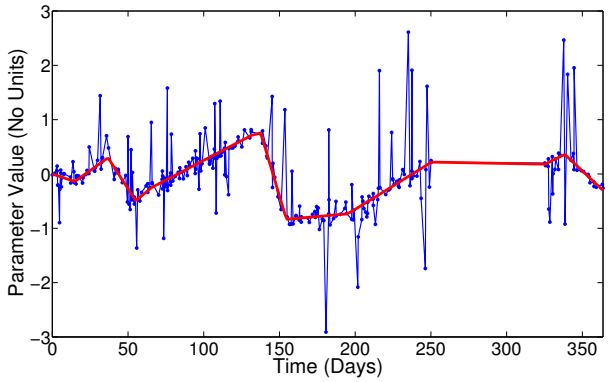

Fig. 1. An example of a simulated signal, with parameters $\sigma_N = 0.1$, $p_N = 0.2$ and $p_o = 0.2$. Also shown is the true underlying piecewise-linear trend signal from which the simulated data is derived.

2) Physiological data: Electrocardiogram, forced spirometry, blood pressure, pulse oximetry and weight measurements were collected from 24 home-dwelling patients using a remote monitoring system called the TeleMedCare Health Monitor - TMC-HM (TeleMedCare Pty. Ltd. Sydney, Australia). The participant ages ranged from 54-92 years and were suffering either chronic obstructive pulmonary disease and/or congestive heart failure. The participants were monitored from February 2007 to January 2008, generating measurements approximately daily. From these recorded measurements, a number of physiological parameters where derived: heart rate (HR), from the electrocardiogram; forced expiratory volume in one second (FEV1), from the forced spirometry; systolic blood pressure (SBP) and diastolic blood pressure (DBP); arterial oxygen saturation (SpO_2), from the pulse oximetry; and finally weight using a standard weight scale. In total, 144 (24 patients $\times$ 6 parameters) were obtained for analysis. Fig. 2 shows a sample plot of physiological data for a subject's weight over approximately one year.

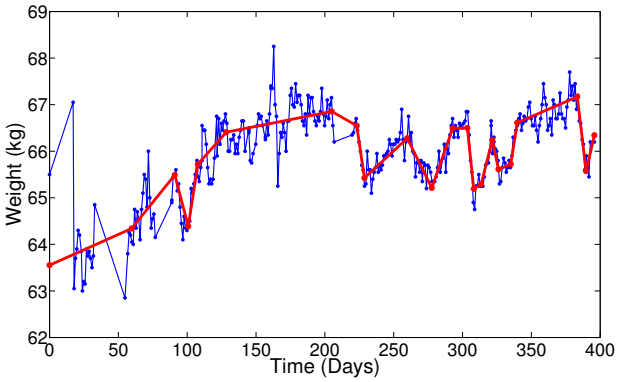

Fig. 2. An sample of approximately one year of weight measurement. Also shown is the human scorers piecewise approximation of these data.

B. Piecewise-linear Regression

A piecewise linear fit [3], $f(x) = w_1 x + w_2 (x - b_1)_+ + w_3 (x - b_2)_+ + \cdots + w_{K+1} (x - b_K)_+ + w_{K+2}$, for the points x_i and y_i, $i \in \{1, ..., N\}$, with breakpoints at $x = b_k$, $k \in \{1, ..., K\}$, where,

$$(x)_+ = \begin{cases} x & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases}, \quad (3)$$

is obtained by solving (1) as $\mathbf{w} = \mathbf{X}^+\mathbf{y}$, by calculating $\mathbf{X}^+$, the pseudoinverse of $\mathbf{X}$.

The difficulty in obtaining a meaningful piecewise regression lies in the selection of the K breakpoint locations, b_k, and in the definition of a cost function which is optimized in the ideal regression. The cost function to be maximized, J, is simply the sum of the squared errors, $J = \sum_{i=1}^{N} (y_i - f(x_i))^2$, subject to the restriction that no error can be greater than some preset limit, e_{max}: $|y_i - f(x_i)| < e_{max} \forall i$.

To select the breakpoints, an exhaustive search of breakpoints is completely infeasible; a heuristic approach is more suitable. A backward selection search, initializing all $b_k = x_k$ for $k \in \{1, ..., N\}$ (i.e. every point is a breakpoint), may

seem achievable; however, the calculation of the pseudoinverse, which involves the inversion of a $(K+2)\times(K+2)$ matrix, becomes problematic for large numbers of breakpoints.

The approach employed here is to limit the number of breakpoints to approximately 50, by using a 56 day sliding window, with an overlap of 28 days. Backward selection involves iteratively removing the breakpoint, from the set of 50 or so possible breakpoints, that gives the minimum increase in J. Breakpoints are sequentially removed until any further removal of breakpoints would violate the maximum single error allowed by e_{max}. From each 28 day window, the remaining breakpoints, in addition to the start and end of the current analysis window, are retained as possible breakpoints for later analysis.

When the scanning process has been completed for all analysis windows, the scan is reiterated, using a variable window size, defined by the span of a maximum of 50 points, until the selection of breakpoints has converged to a fixed number.

C. Selecting e_{max}

The distinction between what variation constitutes noise, or natural variation, and what represents a significant change in the underlying signal, is highly dependent upon the time scale, Δt, which is of interest. In this paper, $\Delta t = 7$ days was chosen.

To select a value for e_{max}, the following procedure is implemented. The variation over the time scale, Δt, is estimated by finding the standard deviation of the data within $\Delta t/2$ days of each point, after linear detrending. Once a standard deviation has been estimated for a window about all data points, the median value is chosen as σ_S.

Using this estimate of the σ_S, any data points whose nearest points on either side, within a Δt day window, both differ from the point in question by more than $3\sigma_S$ are removed. Since these outliers will usually produce an overestimation of σ_S, it is re-estimated, as before, for use later.

The long-term variation in the signal, σ_L, is now calculated as the standard deviation of the entire signal, with the previously detected outliers removed. The choice of the maximum error for the piecewise-linear fit is set at $e_{max} = \sigma_L/4$.

D. Data Smoothing

After the initial outlier removal, described in section II-C, data smoothing was performed using robust locally weighted scatterplot smoothing (robust LOWESS) [4]. Traditional LOWESS regression replaces a data point by a value obtained from a weighted linear fit of a number of data points within a region, termed the 'span', about the point in question. The weighting assigned to each point in the linear fit, within the span, is given as a non-linear function of their difference along the x ordinate, with a larger distance corresponding to a lesser weighting. The robust variation of LOWESS regression incorporates median filtering to remove extreme outlying data points, such that the smoothing is not grossly skewed by these outliers.

The choice of span determines the amount of smoothing performed on the data. In turn, this is a function of the ratio between the inherent variability, σ_S, over the time-scale, Δt, in question (7 days) and the long-term variability in the underlying signal, σ_L. The heuristic employed to select a span, s, is given by: $s = \left(1 + \frac{3\sigma_S}{\sigma_L}\right)\Delta t$.

In case of a poor signal-to-noise ratio, $\sigma_S/\sigma_L \to 1$, giving a maximum span of $4\Delta t$ (approximately 28 days). For a high signal-to-noise ratio, $s \approx \Delta t$.

The piecewise-linear breakpoint search procedure is applied to this smoothed signal, using the worst case error limit e_{max}.

E. Human Scoring

A graphical user interface (GUI), developed in MATLAB, was used to allow human scorers to manually fit a piecewise-linear curve to the data. The scorers were free to add as many breakpoints as they preferred, but where briefed to only use the minimum number necessary to adequately fit the data. The GUI provides the option to insert, delete or move breakpoint locations. Three scorings were obtained from separate individuals, for each of the 125 simulated signals. Only one scoring was obtained for each of the 144 physiological data signals.

F. Performance Metrics

The performance metrics for the simulated data include the root-mean-squared error (RMSE) value of the difference between the true underlying signal and the piecewise-linear fit, as performed by the algorithm described in section II-B and each of the three scorers – noting that the simulated data typically spans the interval $[-1, +1]$.

Since it is impossible to know what the true underlying signal is for the physiological data, the single human scoring is used as the 'gold-standard' estimation. Performance metrics for the physiological data include the RMSE difference between the piecewise-linear fit determined by the algorithm and the human 'gold-standard' scoring. Also, here we present an estimate of the variance of the signal over a short-term window (7 days), termed σ_S, and over the entire long-term duration of the signal, termed σ_L. In addition the RMSE is listed as a fraction of the values σ_S and σ_L.

III. RESULTS

Table I, shows the mean, μ, and standard deviation, σ, of the RMSE between each true underlying signal and the piecewise-linear fit to the simulated data, for each of the three human scorers and the proposed algorithm. The results are grouped according to the amount of noise, σ_N, applied to the signal, noting that there are varying percentages of outlying points and data outages for the 25 signals within each σ_N category.

Table II shows the comparison between the human and algorithmic fits to the real physiological longitudinal data. Shown is the mean, μ, and standard deviation, σ, averaged

TABLE I

A SUMMARY OF THE PERFORMANCE RESULTS FOR SIMULATED DATA.

	Scorer 1	Scorer 2	Scorer 3	Algorithm
σ_N	$\mu \pm \sigma$	$\mu \pm \sigma$	$\mu \pm \sigma$	$\mu \pm \sigma$
0.01	0.01 ± 0.004	0.02 ± 0.018	0.01 ± 0.008	0.03 ± 0.016
0.05	0.03 ± 0.013	0.04 ± 0.012	0.04 ± 0.046	0.03 ± 0.023
0.1	0.05 ± 0.012	0.07 ± 0.019	0.05 ± 0.020	0.05 ± 0.022
0.5	0.19 ± 0.055	0.28 ± 0.076	0.27 ± 0.067	0.17 ± 0.025
1.0	0.35 ± 0.109	0.52 ± 0.164	0.49 ± 0.159	0.32 ± 0.056

TABLE II

A SUMMARY OF THE DIFFERENCE BETWEEN THE ALGORITHMIC FIT OF THE DATA AND THE HUMAN EQUIVALENT.

	σ_S	σ_L	RMSE	RMSE/σ_S	RMSE/σ_L
Parameter	$\mu \pm \sigma$	$\mu \pm \sigma$	$\mu \pm \sigma$	$\mu \pm \sigma$	$\mu \pm \sigma$
HR (BPM)	7.13 ± 4.93	11.50 ± 5.23	2.76 ± 0.77	0.45 ± 0.16	0.27 ± 0.10
FEV1 (L)	0.05 ± 0.02	0.11 ± 0.07	0.04 ± 0.03	0.71 ± 0.57	0.32 ± 0.19
SBP (mmHg)	8.04 ± 1.95	12.84 ± 3.18	4.13 ± 2.05	0.53 ± 0.29	0.32 ± 0.13
DBP (mmHg)	5.41 ± 1.60	10.16 ± 3.37	3.62 ± 2.83	0.73 ± 0.67	0.34 ± 0.16
SpO_2 (%)	1.03 ± 0.54	1.67 ± 1.01	0.51 ± 0.35	0.50 ± 0.16	0.30 ± 0.06
Weight (kg)	0.31 ± 0.09	1.72 ± 1.24	0.49 ± 0.94	1.65 ± 3.19	0.26 ± 0.17

across all 24 subjects. To place the values reported in context for the natural variations of the signals, the estimated values of σ_S and σ_L are also shown.

Fig. 3 illustrates a comparison between the human scoring of the weight data against the piecewise-linear fit derived by the proposed algorithm. This is the same data as shown in Fig. 2, without linear interpolation between the raw data points.

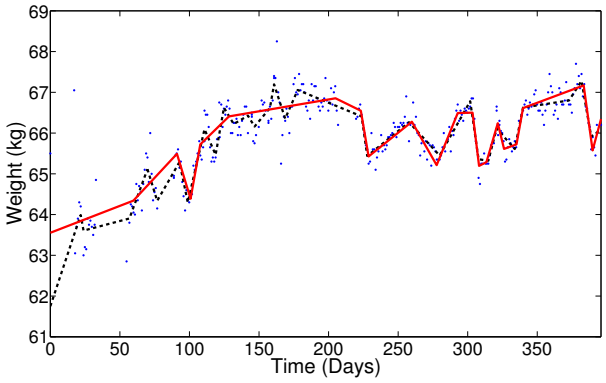

Fig. 3. A sample of approximately one year of weight measurement, identical to that shown in Fig. 2. The 'solid' line shows again the human observer's piecewise-linear fit to the data. The 'dashed' line illustrates the algorithmically determined fit of the data.

IV. DISCUSSION & CONCLUSION

An approach has been described for obtaining a piecewise-linear fit, to longitudinal physiological trend data, comparable with a similar fitting performed by a human observer, using a GUI. The technique has been applied to both simulated and real data, and a comparison performed against the human scoring for each.

From the μ values in Table I, it is evident that the algorithm on average performs equally well, or better than all human scorers; with the exception of the case where $\sigma_N = 0.01$, for which errors were observed to occur mainly around the breakpoints of the true underlying piecewise-linear trend due to the alteration of the signal caused by the smoothing of the LOWESS regression; although the magnitude of the errors, as a fraction of the dynamic range of the signal (± 1) is insignificant. The greatest improvement over the human scoring was achieved on the noisiest data ($\sigma_N = 1.0$).

Table II demonstrates that the algorithm was also able to match the human scoring of the physiological data reasonably well. All RMSE/σ_L values fall below 0.35, indicating that as a fraction of the long-term variation of the signal, the human scoring is tracked well. Indeed, this is echoed with similar results in the RMSE/σ_S ratio; except for the weight parameter, which gives RMSE/$\sigma_S = 1.65$. This is perhaps indicative that the heuristic selection of $e_{max} = \sigma_L/4$ is not the optimum choice.

Very few works relating to the interpretation of longitudinal home-telecare data have been described, possibly due to the relative newness of such data. Bellazzi, *et al.* [5] have devised a more abstract technique for analyzing data from diabetic patients, but their method does not attempt to characterize the data in a detailed manner, which is still open to intuitive interpretation by a clinician.

Future work will focus on the identification of which trends should be considered significant in a clinical context and exactly how these trends may map to an indicator of subject health status and enhance the utility of a decision support system for telecare monitoring.

REFERENCES

[1] B. Celler, N. Lovell, and D. Chan, "The potential impact of home telecare on clinical practice," *Medical Journal of Australia*, vol. 171, pp. 512–521, 1999.

[2] J. Basilakis, N. Lovell, and B. Celler, "A decision support architecture for telecare patient management of chronic and complex disease," in *Proc. 29th Annual International Conference of the IEEE EMBS*, Lyon, France, 2007, pp. 4335–4338.

[3] D. S. Paulson, *Handbook of regression and modeling: applications for the clinical and pharmaceutical industries*. CRC Press, 2006.

[4] W. S. Cleveland, "Robust locally weighted regression and smoothing scatterplots," *Journal of the American Statistical Association*, vol. 74, no. 368, pp. 829–836, 1979.

[5] R. Bellazzi, C. Larizza, and A. Riva, "Temporal abstractions for interpreting diabetic patients monitoring data," *Intelligent Data Analysis*, vol. 2, no. 1-4, pp. 97–122, 1998.