PROTOTYPICAL BIOLOGICAL SIGNALS

Zeeshan Syed, John Guttag Massachusetts Institute of Technology

ABSTRACT

In this paper we present a mechanism to construct prototypes of biological activity by robustly aggregating multiple noise-corrupted and time-warped observations of patient data. This allows for a compact representation of clinical activity while improving the quality of information.

Our prototype construction relies on morphology-based sequence-alignment to relate activity across a set of observations and hierarchical aggregation techniques to merge information across signals in an efficient bottom-up manner. The resulting prototype represents an averaged representation of the original observations along both the time and amplitude axes.

When applied to synthetic ECG, the use of prototypes was able to reconstruct the original signal with an error of less than 6% from data corrupted with unit signal-to-noise ratio and time-warping corresponding to rate variations of 5 beats per minute. In a different experiment on a set of acoustic cardiac recordings, the use of prototypes was also able to highlight differences in patients with and without mitral regurgitation.

Index Terms— prototypes, noise reduction, aggregation, signal averaging, biomedical signals

1. INTRODUCTION

Many biological signals exhibit quasi-periodicity, and this property encourages the use of a variety of aggregation techniques to create composite signals. Signal averaging is commonly employed to reduce additive noise and improve the quality of data. An associated advantage of this approach is that it can help address the problem of "information overload" by compressing the overwhelming amounts of patient data collected during long-term monitoring.

Automated information fusion techniques and aggregation operations have been previously applied to a number of different fields. Ensemble averaging is frequently employed, and prototype construction in numerical, ordinal and categorical data has been explored using the plurality rule, medians, Sugeno integrals and ordinal weighted means [1]. The idea of weighted averaging [2] appears in various forms, using possibility theory [3], the minimum energy principle [4], signal-to-noise maximization with the Rayleigh quotient and generalized eigenvalue problem [5], and adaptive weight estimation [6]. In [7], a general discussion is provided on the application of averaging to quasi-periodic biological signals. Combination in this case proceeds by

means of weighted averaging, where each observation influences the composite signal according to its weight and aggregation can be viewed as an optimization problem over the vector of weights using an objective function that measures the distance between the resulting prototype and each observation.

The works described above typically model all interperiod variations across the signals as noise. We relax this view and attempt to separate noise from naturally occurring variations along the time axis in biological signals. This addresses the case of changes across signals resulting from misalignments of activity, rather than corruption of the data. To achieve this, we supplement previous aggregation algorithms with a mechanism to relate time-warped information across observations and combine data hierarchically. This ensures that the averaging process combines consistent physiological activity. Under the timewarped model, we view the construction of prototypes as the merging of information along both the amplitude and time axes, i.e., the goal of our work is to produce representative signals where events have average amplitude and occur at average displacements from some fiducial point.

The structure of this paper is as follows. Section 2 describes both the process of aligning information across a pair of observations and hierarchical aggregation to create a prototype. An evaluation of this work is presented in Section 3. Finally, we conclude with a summary in Section 4.

2. PROTOTYPE CONSTRUCTION

The classic dynamic time-warping algorithm [8] produces the optimal alignment of two sequences u[n] and v[n], each of length N. This alignment can be denoted by $\varphi = (\varphi_u, \varphi_v)$ and represents a mapping of length K between the samples of the two sequences, i.e.:

$$u[\varphi_u(i)] \leftrightarrow v[\varphi_v(i)], \ 1 \le i \le K \tag{1}$$

The process of aggregating u[n] and v[n] then proceeds by decomposing the optimal alignment φ into α , β and γ segments as shown in Figure 1. An α -segment corresponds to a locally maximal (i.e., non-extendible) segment of the alignment path such that multiple samples from u[n] are aligned against a single sample from v[n]. Conversely, a β segment comprises a locally maximal segment of the alignment path where multiple samples from v[n] are aligned against u[n]. The remainder of φ corresponds to γ -segments.

Aggregation proceeds by iterating over all the



Fig. 1. Decomposition of φ into α , β and γ segments: Horizontal runs along the DTW distance matrix correspond to α -segments, vertical ones to β -segments, and diagonal traversals to γ -segments.

decomposed segments in φ . Each α -segment of length l starting at position k (denoted by α_k) first undergoes a twopoint average for $1 \le n \le \lfloor l/2 \rfloor$:

$$\alpha_{k}^{(1)}[n] = \frac{u[\varphi_{u}(k+2n-1)] + u[\varphi_{u}(k+2n)]}{4}$$
(2)

With:

$$\alpha_k^{(1)}[\lfloor l/2 \rfloor + 1] = \frac{u[\varphi_u(k+l-1)]}{2}$$
(3)

Information is then merged with *v* for $1 \le n \le \lfloor l/2 \rfloor + 1$ by:

$$\alpha_{k}^{(2)}[n] = \alpha_{k}^{(1)}[n] + \frac{\nu[\varphi_{\nu}(k)]}{2}$$
(4)

β-segments are handled in an analogous manner to yield $\beta_k^{(2)}[n]$. For γ-segments, there is no compacting or internal $\gamma_k^{(1)}[n]$ state. Instead, for $1 \le n \le l$:

$$\gamma_{k}^{(2)}[n] = \frac{u[\varphi_{u}(k+n-1)] + v[\varphi_{v}(k+n-1)]}{2} \quad (5)$$

The algorithm for aggregation cycles through all segments, and concatenates the respective $\alpha_k^{(2)}$, $\beta_k^{(2)}$ and $\gamma_k^{(2)}$ sub-signals together to yield an aggregate signal x[n]. For alternate α or β segments with even length (no distinction is drawn between the two cases), the last element corresponding to (4) is left out.

It is critical that this pair-wise aggregation of observations yields a signal of the same length as the ones being combined. This conservation of length property permits the hierarchical approach to compacting activity across multiple observations described at the end of this section.

<u>Theorem</u>: For two inputs u[n] and v[n] of length N, the x[n] produced by this algorithm is also of length N.

<u>Proof</u>: The total length of the aggregated sequence x[n] can be found by distinguishing between α and β segments of even or odd length. α_{odd} and β_{odd} segments of odd length l merge together l + 1 samples across both signals (l samples along one signal, and one sample along the other) to yield $\alpha^{(2)}$ or $\beta^{(2)}$ sub-signals of length $\lfloor l/2 \rfloor + 1$. Since l is odd:

$$\left\lfloor \frac{l}{2} \right\rfloor + 1 = \frac{l+1}{2} \tag{6}$$

Each α_{odd} or β_{odd} segment therefore reduces the number of distinct samples along it across both signals by half. A similar effect can be seen for γ -segments, each of which compact 2*l* distinct samples into a $\gamma^{(2)}$ signal of length *l*. Now, all that remains is to prove that α_{even} and β_{even} segments also compact the number of distinct samples along them across both signals by half.

In the case of α_{even} and β_{even} segments, the corresponding lengths of $\alpha^{(2)}$ and $\beta^{(2)}$ are alternately $\lfloor l/2 \rfloor + 1$ and $\lfloor l/2 \rfloor$. This follows from the fact that as described earlier, for alternate α or β segments with even length, the last element corresponding to (4) is left out. The resulting effect due to this approach is that every odd occurrence of an α_{even} or β_{even} segment has a length exceeding (l + 1)/2 by half a sample, while every even occurrence has a length that is less by the same amount, which counterbalances the earlier excess. In other words, the net effect of having an even number of α_{even} and β_{even} segments is that cumulatively, these segments compact samples along both signals by half.

We can prove the existence of an even number of α_{even} and β_{even} segments as follows. Each DTW alignment path starting at (1,1) and ending at (*N*,*N*) can be described through 2N - 1 movements in the \rightarrow or \downarrow direction. Denoting the *m*-th α , β and γ segments by α_m , β_m and γ_m and length by the operator *L*(.), it is easy to verify that:

$$2N-1 = \sum_{m_{\alpha}} L(\alpha_{m_{\alpha}}) + \sum_{m_{\beta}} L(\beta_{m_{\beta}}) + 2\sum_{m_{\gamma}} L(\gamma_{m_{\gamma}}) + m_{\alpha} + m_{\beta} - 1$$
(7)

Replacing the summation terms by w_{α} , w_{β} and w_{γ} :

$$2N - 1 = w_{\alpha} + w_{\beta} + 2w_{\gamma} + m_{\alpha} + m_{\beta} - 1$$
 (8)

Distinguishing between even (e) and odd (o) length terms this can be rewritten as:

$$2N = w_{\alpha,e} + w_{\alpha,o} + w_{\beta,e} + w_{\beta,o} + 2w_{\gamma} + m_{\alpha,e} + m_{\alpha,o} + m_{\beta,e} + m_{\beta,o}$$
(9)

The terms corresponding to $2w_{\gamma}$, $w_{a,e}$ and $w_{\beta,e}$ must always be even. Removing them:

$$2N' = w_{\alpha,o} + w_{\beta,o} + m_{\alpha,e} + m_{\alpha,o} + m_{\beta,e} + m_{\beta,o} \quad (10)$$

 $w_{a,o}$ can only be odd if an odd number of α_{odd} -segments are present, i.e., $m_{a,o}$ is odd. Similarly, the condition for $w_{\beta,o}$ to be odd is that $m_{\beta,e}$ is odd. From this follows:

$$w_{\alpha,o} + w_{\beta,o} \pmod{2} = m_{\alpha,o} + m_{\beta,o} \pmod{2}$$
 (11)



Fig. 2. Hierarchical aggregation topology: The complete binary tree representation corresponds to a flat organization of observations that are aggregated in a breadth-first manner in pairs.

This means that:

$$2N'' = m_{\alpha e} + m_{\beta e} \tag{12}$$

An even number of α_{even} and β_{even} segments must therefore exist. \blacksquare

The combined effect of the halving associated with α_{odd} , β_{odd} , γ , α_{even} and β_{even} segments is that a total of 2N samples across the two signals combined is aggregated to yield a composite sequence of length N. The net effect of the pairwise aggregation process is therefore to combine information in both amplitude and time, yielding a sequence that has the same length as the initial observations.

Figure 2 illustrates how a complete binary tree topology with depth log(M) can be used to merge information in a uniform, unbiased manner. This contains O(M) internal nodes, each corresponding to an $O(N^2)$ aggregation step associated with DTW-based pair-wise aggregation.

For very large observation sets, random sampling can be used to reduce M, or observations can be down sampled to obtain a quadratic speedup associated with the $O(N^2)$ factor.

3. EVALUATION

We evaluated our techniques by applying them to synthetic ECG data and to acoustical cardiac data acquired using an electronic stethoscope.

3.1. Synthetic ECG

We used the synthetic ECG generator of [9] to evaluate the ability of our techniques to recover an original signal from observations corrupted by varying levels of noise and timewarping. Noise was specified directly to the generator in millivolts, while time-warping was expressed as the standard deviation of heart rate in terms of beats per minute. Since the number of observations generated was finite, the timewarping was not always centered around the original signal, i.e, the warped beats could on average be longer or shorter.

We tested our algorithm under various test conditions and calculated the root mean-square error (RMSE), maximal absolute difference (MD), noise-to-signal ratio (NSR), and dynamic time-warping cost (DTWC) between the aggregate prototype signals and the original signal. 100 observations were used, each corresponding to ECG sampled at 256 Hz with a mean heart rate of 80 beats per minute. The position of the R wave was used as the fiducial point.

The penalty under RMSE and NSR for deleting a relevant event can be half that for detecting the event but repositioning it at a location slightly displaced from the original. This occurred in our experiments because of the finite sample issue discussed at the start of this subsection. Similarly, the MD measure penalizes deletion as much as displacement. Since missing events altogether is typically more serious in the clinical setting than minor relocations, we consider the DTWC, which penalizes deletion more than displacement, to be a more useful measure of error.

The results of these experiments are shown in Figure 3. The left column of the figure shows the error using each of the four measures between the ensemble average and the original signal. The right column shows the corresponding errors for the prototype. In the absence of time-warping, the ensemble provides a least squares optimal estimate of the original signal. Even so, there is not much difference between the performance of the ensemble and that of the prototype in the absence of time-warping. Furthermore, with even small amounts of time-warping the prototype provides a significant performance improvement.



x-axis: noise-to-signal ratio of generated observation: y-axis: time-warping in beats per minute

Fig. 3. Synthetic ECG error: From top to bottom, the RMSE, MD, NSR and DTWC of the ensemble average and prototype relative to the deterministic ECG signals are shown for different additive noise and warping levels (each cell corresponds to a different randomly generated set of observations).



Fig. 4. Normal heart: An individual beat (top) is compared against the prototynical beat (bottom) in three frequency bands.



Fig. 5. Mitral regurgitation: An individual beat (top) is compared against the prototypical beat (bottom) in three frequency bands.

3.2. Acoustic Cardiac Data

In another work [10], we studied the use of prototypical signals to improve the accuracy of automated diagnosis of mitral regurgitation (MR) from acoustical recordings. We used the same classifier on individual beats from the original signal and on prototypical beats. When applied to a dataset comprising 11 patients with MR and 28 with normal hearts

or benign murmurs, the use of a prototype was able to reduce false positives from nine to three. The number of false negatives was reduced from two to one.

The greater false positive rate while analyzing individual beats resulted from noise or natural variations that mimicked MR. Similarly, variations in individual beats can obscure a trend that does exist. The prototypical calculation removed these variations while retaining information that recurred across the examination.

The overall effect of the use of the prototype was to improve the quality of the information available to the classifier. Variations that could lead to erroneous diseases classifications were removed, while making cases where an underlying disorder existed more apparent. Figures 4 and 5 illustrate this effect by comparing individual and prototypical beats for patients with MR and normal hearts. For the normal heart, the prototype is able to remove variations and show the presence of peaks at S1 and S2.. In the case of the patient with MR, the presence of the recurring mid-systolic clicks and the high frequency murmur can be seen clearly in the prototype.

4. SUMMARY

In this paper we presented a mechanism that uses sequence alignment and hierarchical aggregation to create prototypical signals that compactly and robustly represent physiological activity in the presence of noise and time-warping. We demonstrated that when applied to a synthetic ECG signal our techniques do a good job of recovering the original signal in the presence of unit NSR and a time-warping error corresponding to five beats/minutes. We also showed how our techniques can be used to provide improved visual representations of a cardiac disorder that has an acoustical signature.

5. REFERENCES

[1] J Domingo-Ferrer et al, "Median-Based Aggregation Operators for Prototype Construction in Ordinal Scales", *International Journal of Intelligent Systems*, 18, pp. 633-655, 2003

[2] R Yager, "On Mean Type Aggregation", IEEE TSMC, 26(2), 1996

[3] D Dubois et al, "Possibility Theory in Information Fusion", Intl. Conf. on Information Fusion, 2000

[4] Z Fan et al, "Weighted Averaging Method for Evoked Potentials: Determination of Weighted Coefficients", *IEEE EMBS Conf.*, 1992

[5] C Davila et al, "Weighted Averaging of Evoked Potentials", *IEEE Trans. on Biomedical Engineering*, 39, 1992

[6] E Bataillou, et al, "Weighted Averaging Using Adaptive Estimation of the Weights", *Signal Processing*, 44, 1995

[7] J Leski, "Robust Weighted Averaging", IEEE TBME, 49, 2002

[8] C Myers, L Rabiner, "A Comparative Study of Several Dynamic Time-Warping Algorithms for Connected Word Recognition", *The Bell System Technical Journal*, 60(7), 1981

[9] P McSharry et al, "A Dynamic Model for Generating Synthetic Electrocardiogram Signals", *IEEE TBME*, 50, 2003

[10] Z Syed et al, "A Framework for the Analysis of Acoustic Cardiac Signals", *to appear, IEEE TBME*