QUANTIFYING MORPHOLOGY CHANGES IN TIME SERIES DATA WITH SKEW

Phil Sung, Zeeshan Syed and John Guttag

Massachusetts Institute of Technology

ABSTRACT

This paper examines strategies to quantify differences in the morphology of time series while accounting for time skew in the observed data. We adapt four measures originally designed for signal shape comparison: Dynamic Time-Warping (DTW), Earth Mover's Distance (EMD), Fréchet Distance (FD), and Hausdorff Distance (HD). These morphology difference metrics on time series are compared in discriminative power and noise resistance on ECG signals as well as on a synthetic dataset. We use data from our experiments to shed light on the relative strengths of the methods.

Index Terms- shape matching, time skew

1. INTRODUCTION

Quantifying the similarity of two time series is of importance in many signal-processing applications. Simple metrics such as Euclidean distance are not suitable when the signals have variable amounts of time skew. In this case, comparing samples by their timing may cause parts of the signals associated with different phenomena to be compared, leading to a poor estimate of similarity.

Methods that relate corresponding parts of two signals before measuring differences are useful in a variety of settings (e.g. [1, 2]). In this paper we consider four morphology difference measures for real-valued time series represented as sequences of time-amplitude pairs, $(t_1, v_1), \ldots, (t_n, v_n)$. Some of these measures are adapted from earlier methods for comparing shapes in a general metric space. We consider Dynamic Time-Warping (DTW) as well as adaptations of Earth Mover's Distance (EMD), Fréchet Distance (FD), and Hausdorff Distance (HD).

We evaluate each metric in two ways. We first examine the distances produced when the metric is used to discriminate one shape from another. We then examine the robustness of the metric when its inputs are corrupted by noise. Both analyses are performed on two datasets: synthetic data from the cylinder-bell-funnel problem [3] and ECG data from the PhysioNet MIT-BIH dataset [4]. Although various classification techniques have previously been applied to identify broad and clinically recognized classes of behavior in ECG data [5], our focus here is on quantifying subtler morphologic differences. One possible application is the measurement of patterns of myocardial instability that may portend high risk [2].

Section 2 summarizes the four metrics and the adaptations we propose to them. Section 3 describes the evaluation procedure and results. Section 4 concludes with a discussion.

2. TECHNIQUES

2.1. Dynamic Time-Warping

The Dynamic Time-Warping (DTW) distance metric [6] computes the distortion needed to align two time series. An *alignment* of two sequences A and B, of length m and n respectively, is a sequence of integer pairs $(\phi_A[1], \phi_B[1]), (\phi_A[2], \phi_B[2]), \dots, (\phi_A[k], \phi_B[k])$ satisfying the boundary conditions $\phi_A[1] = 1, \phi_B[1] = 1, \phi_A[k] = m$, and $\phi_B[k] = n$, as well as these continuity conditions:

$$\phi_A[j] \le \phi_A[j+1] \le \phi_A[j] + 1, \phi_B[j] \le \phi_B[j+1] \le \phi_B[j] + 1.$$

Intuitively, each ordered pair $(\phi_A[i], \phi_B[i])$ matches two elements to be aligned. The DTW distance between A and B is the minimum sum-of-squares difference over all allowable alignments:

$$DTW(A, B) \equiv \min_{\phi_A, \phi_B} \sum_{i} (A[\phi_A[i]] - B[\phi_B[i]])^2.$$
(1)

DTW captures both amplitude and timing differences between the signals. Timing differences are captured by an increase in the length k of the alignment (i.e. a larger number of summed terms).

2.2. Earth Mover's Distance

Earth Mover's Distance (EMD) [7] is a metric for comparing two non-negative signals based on the amount of *work* needed to construct one signal by deforming the other (we obtain non-negativity by taking the absolute value of each signal, if necessary). The signal with larger total mass is considered to be the *source* (A) and the smaller signal is considered to be the *goal* (B).

The goal signal is constructed by moving mass from where it exists in the source signal. Suppose $f(t_A, t_B)$ is the amount of mass to be moved from point t_A to point t_B . The function f describes an allowable movement if it does not exhaust the mass available at the source and satisfies the demand at the goal:

$$A[t] - \sum_{t_B} f(t, t_B) \ge 0 \quad \forall t$$

$$A[t] + \sum_{t_A} f(t_A, t) \ge B[t] \quad \forall t.$$
(2)

EMD is the minimum amount of work needed (over all allowable functions f) to construct the goal, where one unit of work is needed to move a unit mass a unit distance. The result is normalized by the total mass of the goal signal:

$$\operatorname{EMD}(A,B) \equiv \frac{\min_{f} \sum_{t_{A},t_{B}} |t_{A} - t_{B}| f(t_{A},t_{B})}{\sum_{t} B[t]}.$$
 (3)

The problem of computing the minimum work can be expressed as a linear program.

2.3. Fréchet Distance

Fréchet Distance (FD) [8] is often described by analogy to a man walking a dog: a man and a dog, connected by a leash, walk along different paths. The man and the dog may move at any speed they wish, but neither may move backwards. FD is the minimum leash length that allows the man and the dog to traverse their paths.

Formally, given two directed paths parameterized by functions A and B mapping [0, 1] to the set S, and a metric $d(s_1, s_2)$ on S, FD is defined as

$$FD(A,B) \equiv \min_{\phi_A,\phi_B} \max_{0 \le t \le 1} d(A(\phi_A(t)), B(\phi_B(t))).$$
(4)

where ϕ_A and ϕ_B (intuitively, the positions of the man and dog as functions of time) may be any monotonically non-decreasing continuous functions from [0,1] to [0,1]satisfying the boundary conditions $\phi_A(0) = 0$, $\phi_A(1) =$ 1, $\phi_B(0) = 0$, and $\phi_B(1) = 1$.

To apply FD to time-series data, we treat a time series as a path in time-amplitude space and define a distance metric of the form $d((t_1, v_1), (t_2, v_2))$.

2.4. Hausdorff Distance

Hausdorff Distance (HD) [9] measures the difference between two closed and bounded sets of points. Given a metric $d(s_1, s_2)$, the distance from a single point a to a set B can be defined as the distance from a to the nearest point in B, i.e. $\min_{b \in B} d(a, b)$. To generalize this notion to two sets A and B, we consider the maximum distance from any point in either set to the other set:

$$\operatorname{HD}(A,B) \equiv \max\left\{ \max_{a \in A} \min_{b \in B} d(a,b), \max_{b \in B} \min_{a \in A} d(a,b) \right\}.$$
(5)

To apply HD to time-series data, we treat each input sample as a point in time-amplitude space (t, v) and define a metric $d((t_1, v_1), (t_2, v_2))$.

3. EVALUATION AND RESULTS

3.1. Selecting scaling parameters for FD and HD

FD and HD are both parameterized by a distance metric $d((t_1, v_1), (t_2, v_2))$. Such a metric must make some trade-off in how it weights changes in time against changes in amplitude. The particular choice of metric may affect performance on applications to which the distance measure is applied. To assess FD and HD, we evaluate their performance with a range of metrics. In particular, we consider metrics of the form

$$d((t_1, v_1), (t_2, v_2)) \equiv \sqrt{(v_2 - v_1)^2 + \alpha^2 (t_2 - t_1)^2}$$
(6)

where α (the "scaling parameter") is a positive number which sets the relative weight of unit changes in time and amplitude.

3.2. Discrimination ability in the cylinder-bell-funnel problem

To evaluate the utility of the various morphologic distance measures for distinguishing between shapes, we compared the distances yielded by each method on a set containing 500 randomly generated instances of each of the three eponymous classes (shapes) of the cylinderbell-funnel (CBF) problem [3].

For each metric, we computed the pairwise distances between all 1500 elements of the set. We then measured the ratio of the average distance in pairs containing different shapes to the average distance in pairs of the same shape. We refer to this ratio as the inter/intra distance ratio (IIDR). High values of the IIDR indicate that the method distinguishes between instances of different classes while ignoring differences between instances of the same class.

For FD and HD, we evaluated the IIDR under various scaling parameters as described in Section 3.1. The results are shown in Figure 1. For subsequent experiments on this dataset, we evaluated FD and HD using the scaling parameters that yielded the highest IIDR here.

The IIDR observed for each of the four methods is reported in Table 1. DTW, followed by EMD, showed the best performance among the methods considered.



Fig. 1. Discrimination performance (IIDR) on CBF data as a function of the scaling parameter.

Method		IIDR
DTW		2.06
EMD		1.82
Fréchet	$\alpha=0.105$	1.38
Hausdorff	$\alpha=0.087$	1.58

Table 1. Discrimination performance (IIDR) on CBFdata for the four morphologic distance metrics.

3.3. Noise resistance in the cylinder-bell-funnel problem

We evaluated the noise resistance of each morphologic distance metric by comparing the morphologic distances obtained from given inputs to the morphologic distances obtained when those inputs were corrupted by noise.

We synthesized 100 pairs of random cylinder-bellfunnel instances (i.e. 200 instances) and computed the morphologic distance between the elements of each pair. We then corrupted each instance with Gaussian white noise of a fixed variance. The computation of morphologic distances was repeated on the noisy data. The correlation coefficient of the two output sequences was used as a measure of how robust the metric was against noise. We evaluated this correlation for various values of noise variance.

A graph showing the morphologic distance correlations observed for various noise levels is presented in Figure 2. DTW showed the highest noise resistance at all noise levels, followed by EMD.

3.4. Discrimination ability on ECG data

We performed experiments analagous to those in Sections 3.2 and 3.3 on electrocardiogram (ECG) data. We acquired ECG data for 23 patients from a subset of the MIT-BIH PhysioNet dataset [4] that is intended to be a representative sample of ECG waveforms. The data were annotated with cardiologist-supplied labels for each heartbeat. We excluded 9 patients who did not have at least two differently labeled beats each of which occurred at least five times. For each of the remaining 14 patients,



Fig. 2. Correlation between morphologic distance measures on clean and noisy CBF instances, as a function of noise applied.



Fig. 3. Discrimination performance (average IIDR) on ECG data as a function of the scaling parameter.

we randomly selected pairs of heartbeats with the same label and pairs of heartbeats with different labels. We then computed the IIDR—here, the ratio of the average distance between beats of different labels to the average distance between beats of the same label. The average IIDR across all patients was considered to be a measure of discrimination performance in this application.

For FD and HD, appropriate scaling parameters were selected as in Section 3.2; the average IIDR as a function of scaling parameter is shown in Figure 3. The average IIDR for each of the four metrics is reported in Table 2. In this application, DTW and EMD again showed the best discrimination performance.

3.5. Noise resistance on ECG data

We evaluated the noise resistance of each method as in Section 3.3, using the pairs of ECG signals from the previous experiment. Figure 4 shows the correlation between the morphologic distances on clean data and the morphologic distances when the data were corrupted by noise, as a function of the noise variance. DTW showed the highest noise resistance at all noise levels, followed

Method		IIDR
DTW		2.66
EMD		2.16
Fréchet	$\alpha=0.0077$	1.74
Hausdorff	$\alpha=0.0077$	1.76

Table 2. Discrimination performance (average IIDR) onECG data for the four morphologic distance metrics.



Fig. 4. Correlation between morphologic distance measures on clean and noisy ECG data, as a function of noise applied.

by HD.

4. CONCLUSION

DTW showed the best performance in all of the experiments, and EMD showed the second-best performance in three of the four experiments. We believe that the superior performance of the DTW and EMD methods is in part due to the fact that they sum a particular cost function across the entire signal, yielding a cumulative assessment of morphology differences.

In contrast, FD and HD look at the input samples to see where some feature is maximized, and then use that as their computed distance. Consequently, those metrics are quite sensitive to differences (including random noise) in limited regions of the two signals—differences that are not necessarily representative of morphologic differences across the entire signal. In many settings, measurements are accompanied by significant amounts of noise, so it is critical that signal processing techniques are robust against this noise.

One additional consideration is that the use of FD and HD (as described here) is complicated somewhat by the need to select an appropriate metric. This choice of metric can have a significant effect on discrimination performance.

We note that our results should not be interpreted as discouraging in general the use of any of the methods considered here. The best method to use in any situation is liable to depend on the specifics of the application and on a physical intuition for the problem. The various methods also likely have varying sensitivity to different kinds of morphologic differences. Further investigation on additional datasets might yield a more complete picture of the relative strengths of each method. Finally, some of these methods operate in general metric spaces and are thus suitable for comparisons of higherdimensional data; however, we did not evaluate such applications here.

5. REFERENCES

- T. Oates, L. Firoiu, and P. R. Cohen, "Clustering Time Series with Hidden Markov Models and Dynamic Time Warping," in *IJCAI*, 1999.
- [2] Z. Syed et al., "Risk-stratification following acute coronary syndromes using a novel electrocardiographic technique to measure variability in morphology," *Computers in Cardiology*, 2008.
- [3] N. Saito, Local feature extraction and its application using a library of bases, Ph.D. thesis, Yale University, 1994.
- [4] A. L. Goldberger et al., "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.
- [5] N. Maglaveras et al., "ECG pattern recognition and classification using non-linear transformations and neural networks: A review," *Int. Journal of Medical Informatics*, vol. 52, no. 1-3, pp. 191–208, 1998.
- [6] C. S. Myers and L. R. Rabiner, "A comparative study of several dynamic time-warping algorithms for connected word recognition," *The Bell System Technical Journal*, vol. 60, no. 7, pp. 1389–1409, 1981.
- [7] Y. Rubner, C. Tomasi, and L. J. Guibas, "A Metric for Distributions with Applications to Image Databases," in *Proc. IEEE ICCV*, January 1998, pp. 59–66.
- [8] B. Aronov et al., "Fréchet distance for curves, revisited," in *ESA'06*, London, UK, 2006, pp. 52–63, Springer-Verlag.
- [9] D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge, "Comparing images using the Hausdorff distance," *IEEE Trans. PAMI*, vol. 15, no. 9, pp. 850– 863, 1993.