Mapping geometry in heart rate data

Gunnar Carlsson, Jeffrey Danciger, Jason Morton

December 12, 2009

Abstract

We introduce a new methodology for the analysis of heart interbeat interval time series based on the geometric/topological analysis of spaces of frequently occurring *motifs* (short patterns in the data). We use these methods to identify interesting differential features in data from three patient groups: healthy, congestive heart failure, and atrial fibrillation.

1 Introduction

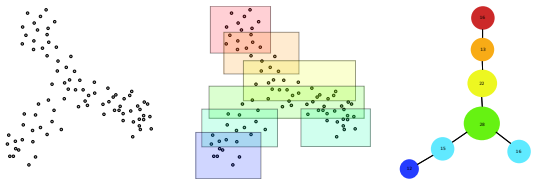
Much effort has been dedicated to better understanding physiologic time series. One common thread in some recent work is the notion that the complexity of the time series yields information about the health of the patient. Costa et. al [2] have developed a complexity measure called multi-scale entropy (MSE) and have successfully applied it to differentiate between patients with three different heart conditions. The fruitful results of this and other complexity measures suggest that there is a great deal of information about the patient contained in a heart interbeat interval time series, perhaps even more so than is thought by the present day medical community. In this article we introduce a new methodology for examining interbeat interval time series that differentiates disease status by identifying specific structural features. The concrete features identified in this article, and with future applications of this method, could provide a useful complement to complexity based features (such as MSE). In particular, we claim that our methodology produces features that may be easier to interpret from a physiologic point of view and therefore could eventually lead to a better understanding of the underlying physiologic systems.

Our methodology is built around two processes. The first is the detection of *motifs*, i.e. recurring patterns of heart rate variation, over multiple time scales that are differentially expressed in patients with differing disease status. The second is the topological analysis of the space of frequently occurring motifs. Our novel method exploits the geometry of the data to produce useful visualizations rather than measurements of mysterious statistics. We note that on the one hand our method produces visualizations that require interpretation, but that on the other hand our method is quite general and can be applied to any time series. As the experiments described here are the first applications of this method, we are not yet able to speculate on the range of data for which the method will be effective.

2 Mapping motifs in interbeat interval sequences

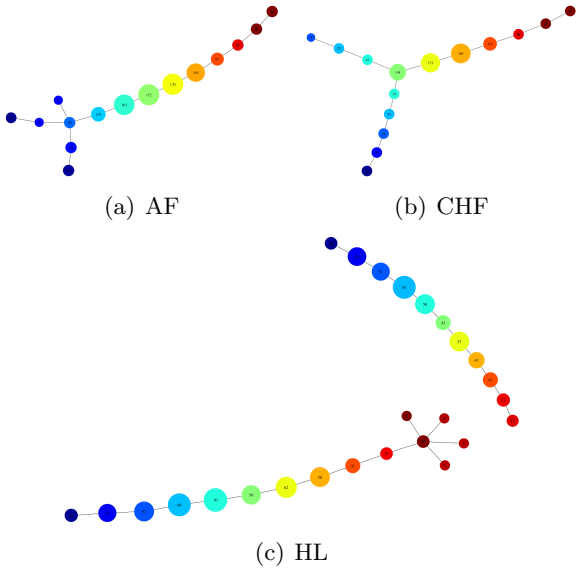
We examined heart interbeat interval sequences collected over 24-hour periods from 72 healthy (HL) patients, 43 patients with congestive heart failure (CHF), and 9 patients with atrial fibrillation (AF). About 3500 N-tuples were selected at random from each time series for varying values of N from 5 to 35. From each such collection, the high variance high density (HVHD) motifs were selected and normalized as follows. The density at each tuple was calculated using a gaussian kernel. The 40% highest variance tuples were selected. From these tuples, the 40% highest density tuples were selected. The remaining tuples were then normalized to have zero mean and

variance equal to one (as we seek features independent of the usual statistical measures). The resulting collection represents the most densely occurring high variance motifs of length N in the time series as a subset of an $N-2$ dimensional sphere centered at the origin in $\mathbb{R}^N$. It is useful to imagine that this discrete set of motifs is sampled from a continuous *motif space*.



(a) data set (b) filter and cluster (c) Mapper output

Figure 1: Schematic of the Mapper algorithm [1]. The data (a) is filtered into overlapping subsets using the height function as filter. Each subset is partitioned into clusters, shown in (b). Mapper outputs a graph (c) capturing geometric features of the data.



(a) AF (b) CHF (c) HL

Figure 2: Results from Mapper with density as filter on the motifs of length $N = 6$ from (a) an atrial fibrillation patient, (b) a congestive heart failure patient, (c) a healthy patient. Note that (c) has two connected components while (a) and (b) have only one.

We used Mapper [1] to explore the geometry of the high variance high density (HVHD) motifs. Mapper is a tool that produces schematic graphs of point cloud data, as in the example depicted in Figure 1. Mapper filters the data into overlapping subsets according to a user defined filter function and then partitions each subset into clusters. Mapper then constructs a graph representing each cluster as a node and connecting two nodes with an edge if the clusters overlap. We used a gaussian kernel density function as the filter in Mapper to produce graph outputs for varying values of N. A sample of the results is shown in Figure 2. These graphs immediately reveal a topological difference between the HVHD motifs of the healthy patient versus those of the other two patients. The motifs from the healthy patient separate into two connected components, while the motifs from both the CHF and AF patient form one large connected component.

In the Mapper analysis of an HL patient, one connected component contains motifs exhibiting a increasing trend (corresponding to decreasing heart rate) while the other component contains motifs exhibiting a decreasing trend (corresponding to increasing heart rate). This phenomenon is illustrated by running Mapper with slope as the filter function (see Figure 3).

2.1 Quadratic Fit Plots

In order to visualize the distribution of motif shapes, we fit each N-tuple to a quadratic $an^2 + bn + c$ and plot the resulting fit parameters b vs. a with color indicating of quality of fit. Roughly, the y-coordinate describes the slope of the motif and the x-coordinate describes the concavity. A simple calculation shows that all points of such a quadratic fit plot must lie inside a *bounding ellipse*, which is given by the fit parameters of normalized N-tuples that are fit perfectly by a quadratic.

Quadratic fit plots for varying motif length N were created for each patient. Although the results vary substantially from patient to patient (see section 3), typical results are shown in Figure 4. These plots demonstrate the varying distribution of motif shapes as we increase the size N of motifs considered. The plots demonstrate very different behavior across patient classes. For small N, the HL motifs separate into two

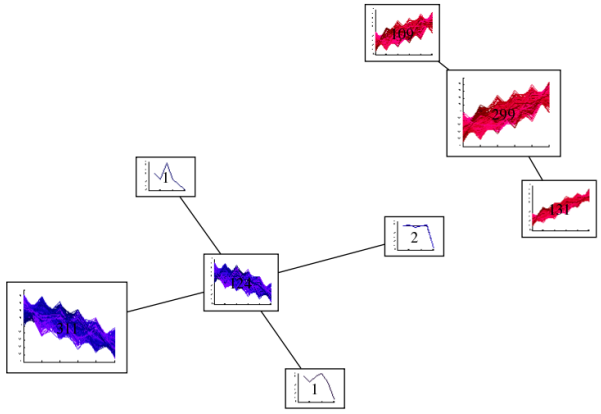
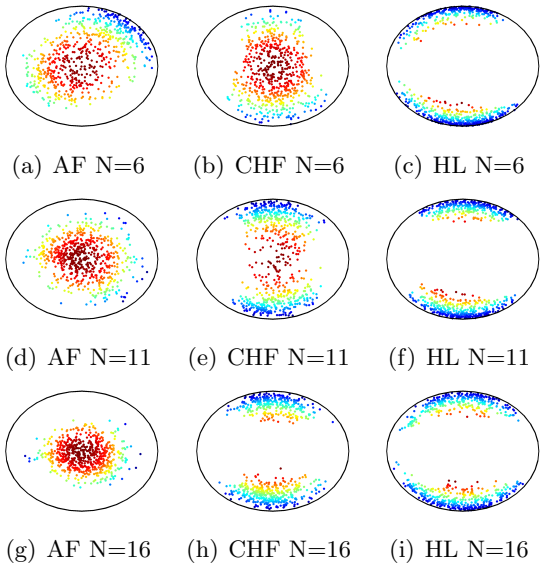


Figure 3: Output from mapper with slope as the filter function on the motifs of length $N = 6$ from an HL patient. Each node depicts the collection of motifs it represents.



(a) AF N=6 (b) CHF N=6 (c) HL N=6

(d) AF N=11 (e) CHF N=11 (f) HL N=11

(g) AF N=16 (h) CHF N=16 (i) HL N=16

Figure 4: Plots of slope vs. concavity obtained from quadratic fits, for motif lengths $N = 6, 11, 16$. Color represents relative error of fit. The bounding ellipse, consisting of the motifs fit perfectly by a quadratic, is depicted in black. For small N, the HL motifs separate into two components (increasing and decreasing), while CHF motifs exhibit no such separation. As N increases the CHF motif space gradually develops separation between increasing and decreasing motifs. Quadratic fit plots of the AF patient motif spaces appear similar to that of Gaussian noise.

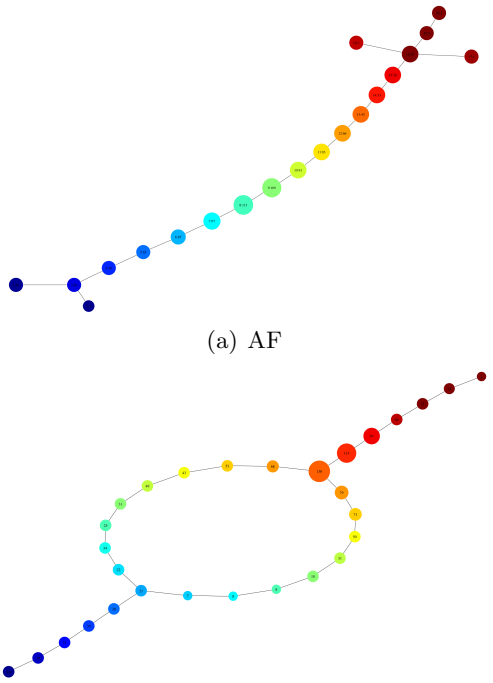
groups (increasing and decreasing), while CHF motifs exhibit no such separation. However, as N increases the CHF motif space gradually develops separation between increasing and decreasing motifs similar to that seen for the HL patient for small N. Quadratic fit plots of the AF patient motif spaces appear similar to those of Gaussian noise.

2.2 Non-quadratic Motifs

The quadratic fit plots described in Section 2.1 reveal a lot about the motif space of an interbeat interval time series. In particular, points falling near the bounding ellipse in these plots are fit very well by a quadratic and the location of such points on the plot therefore accurately describes the shape of the corresponding motif. However, points close to the center of the ellipse have high fit error meaning that quadratic functions fail to fully describe the shape of these motifs. As can be seen in the quadratic fit plots from Figure 4, both CHF and AF patients exhibit many of these *non-quadratic* motifs. We used Mapper to investigate the space of such motifs. Figure 5 displays Mapper outputs from an AF patient and a CHF patient. Note the appearance of a cycle in the CHF patient's Mapper output. This indicates the presence of a nontrivial circle in this patient's non-quadratic motif space. Further analysis reveals that the CHF patient's non-quadratic motif space is made up of oscillatory patterns having similar frequency but having varying phase. On the other hand, the AF patient's non-quadratic motif space seems to lack interesting topological structure.

3 Predicting Disease Status

The quadratic fit plots described in Section 2.1 and Figure 4 depict compelling visual differences between the three patient types. We converted some information from these plots into numerical features as follows. For each of the motif lengths $N = 5, 7, 10, 12, 18, 25, 35$ we calculated the total error of quadratic fit over all of a patient's HVHD motifs. We refer to these 7 features as quadratic fit error (QFE). We also calculated the multi-scale entropy (MSE), as described in [2], at 4 time scales, as well as the mean and variance



(a) AF

(b) CHF

Figure 5: Mapper outputs, with density as filter, from the non-quadratic motifs of (a) an AF patient, and (b) a CHF patient. The presence of a cycle in (b) indicates the presence of a non-trivial circle in the non-quadratic motif space of the patient. This circle corresponds to oscillatory motifs of similar frequency but of varying phases.

(MV) of each time series. In order to test the significance of our QFE features we used various combinations of the three feature sets to classify CHF vs. HL. For each combination of feature sets, we built a simple logistic regression classifier and performed a series of cross validation tests. The results are shown in Table 3. We note that it was sometimes necessary to reduce the number of features using PCA because of the limited number of examples available.

Though the QFE features alone do not classify extremely well, our results show that the QFE features provide a significant complement to the MSE and MV features. We note that any features produced by our method should automatically be independent of mean and variance due to our normalization procedure (Section 2). Due to the limited size of our AF data set, we did not run extensive cross validation tests of classification of AF vs HL or AF vs CHF. We do note, however, that a SVM built with only our QFE features distinguished AF from HL with success rate 99% in a leave one out cross validation test. A similar success rate was achieved using only the mean and variance features.

Features Used	Success Rate	Std Error#
QFE	81.4%	0.3%
MSE	81.3	0.3%
MV	90.1%	0.2%
QFE, MSE	84.6	0.3%
QFE, MV	91.1%	0.2%
MSE, MV	91.1%	0.2%
QFE, MSE, MV	92.1%	0.2%

Table 1: Linear classifiers of CHF vs HL were built using various combinations of three feature sets: Quadratic fit error (QFE), multi-scale entropy (MSE), and mean and variance (MV). The results of a cross validation test are shown here.

References

[1] G. Singh, F. Memoli and G. Carlsson. *Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition.* Point Based Graphics 2007, Prague, September 2007.

[2] M. Costa, A. Goldberger, C.-K. Peng. *Multiscale entropy analysis of complex physiologic time series.* Physics Review Letters Vol. 89 No. 6, August 5, 2002.

[3] Graph images made using *Graphviz.*