Electrocardiogram Signal Classification Based on Fractal Features

AN Esgiar, PK Chakravorty

University of Al Tahadi, Sirte, Libya

Abstract

Atrial fibrillation ECG signals have been classified with fractal features only. The fractal features - fractal dimension, mass dimension and lacunarities were estimated by a new box counting algorithm; called the True Box Counting method. The classification result and stepwise discriminant analysis for these fractal features were determined. It was seen that lacunarities based on higher mass moments were more important than fractal dimension and mass dimension. The results suggest further investigation of lacunarity features.

1. Introduction

The fractal nature of ECG signals has been investigated by many investigators. Fractal dimension, configuration entropy, different moments of the mass distribution, lacunarities (based on different mass moments) etc., are important signatures of a fractal set. Lacunarities are measures of structure in a fractal set (or distribution of gaps in a point set) and are important parameters and if these are used along with fractal dimension then it is expected that more information can be extracted from a fractal set [1].

We have attempted to classify the ECG signals purely on the basis of fractal dimension, mass dimension and lacunarities estimated by a new box counting algorithm, we call this algorithm True Box Counting (TBC) Method. We report here the TBC method, with results of the application of the TBC method to Physionet challenge 2004 ECG signals and discuss briefly the validity of the TBC method. No attempt has been made to extract additional features from fractal dimension, mass dimension and lacunarities.

2. Methods

The atrial fibrillation ECG signals data set files with their annotation files were obtained from Physionet challenge 2004 [2]. This data set contains ECG signals for Training Data and Test Data. For each person (i.e., a case) two signals had been collected. We called them signal1 and signal2. The Training set has 20 cases with 10 N-type (non-terminating AF) and 10 T-type (AF that terminates immediately) cases. The Test set has an unknown mixture of 30 cases belonging to N-type and T-type cases. The ADC (Analogue to Digital Converter) output values essentially integers, both positive and negative are obtained from these ECG signal by using the WFDB package [3] of Physionet tool kit. These constitute the point set in E^2 space as shown in Fig.1. These values i.e., sample numbers and corresponding ADC outputs (x,y) were shifted to the positive quadrant to avoid the negative and zero ADC output values as shown in Fig.2.

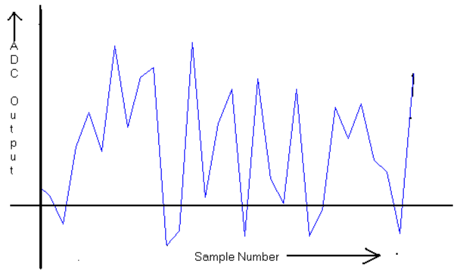

Figure 1. Plot of ADC output Vs. Sample Number.

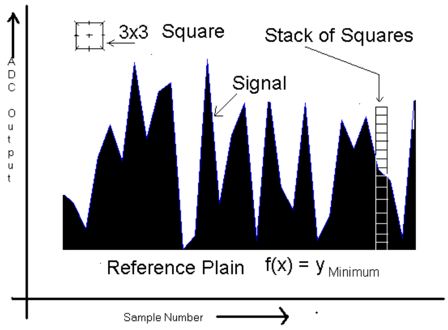

Figure 2. The signal profile is shifted to positive quadrant and a plain profile is constructed from signal profile

2.1. TBC algorithm

A fractal set y = f(x) constituted a point set in E^2 space and since it was translation invariant, we shifted it to the positive quadrant. A line parallel to the abscissa was drawn at $y_{minimum}$. This line was the reference line. All the points in between the signal profile and reference line constituted a plain profile as shown in Fig.2. This plain profile had three straight edges and a jagged edge, which was the signal profile. To calculate different fractal features of this plain object, a stack of vertical square boxes of (3,5,7 ... pixel length) was placed on a point on the reference line of the reconstructed plain profile as shown in Fig.2. A sufficient border margin on the vertical sides of this plain profile was kept while placing these stack of square boxes.

The number of boxes required to span the entire plain profile, $N_{box}(L)$ and different mass-moments $M^q_{box}(L))$ were calculated for different box size (L = 3,5,7 ...) as proposed by Voss[4] yielding an estimate of fractal dimension, mass dimension and lacunarities.

Lacunarity based on second mass moments was estimated as proposed by Mandelbrot[4, 1].

- Lacunarity based on second mass moment:

$$\Delta(L)_2 = \frac{M^2_{box}(L)}{(M_{box}(L))^2} - 1 \qquad (1)$$

Similarly, lacunarity based on higher mass moments was evaluated.

- Lacunarity based on 3^{rd} mass moment:

$$\Delta(L)_3 = \frac{M^3_{box}(L)}{(M_{box}(L))^3} - 3 * \frac{M^2_{box}(L)}{(M_{box}(L))^2} + 2 \qquad (2)$$

For perfectly fractal sets like Cantor set, Sierpinski triangle, Sierpinski curve the lacunarity plot ($\Delta(L)_n$ Vs. L) yields a line parallel to abscissa. For a statistically fractal set, lacunarity $\Delta(L)_n$ is a function of box size, 'L'. For such a curve, we used the power law to best fit the data points.

- The equation for best fit:

$$\Delta(L)_n = (a_n) * (L)^{m_n} - 1 \qquad (3)$$

For perfectly fractal sets the value of index m_n is zero and coefficient a_n represented the lacunarity of the set.

2.2. Classification of ECG signals

True Box Counting (TBC) algorithm was applied on the ADC output of each ECG signal and fractal dimension, mass dimension (first moment of mass) and lacunarities up to 8^{th} moment of mass (a_n and m_n for $\Delta(L)_2$ to $\Delta(L)_8$) were estimated. The decision to choose lacunarities up to 8^{th} moment of mass was arbitrary. For classification, KNN classifier [5] was used and each signal from Test Set was classified and the classification ('N' type or 'T' type with probability value) was obtained. For each individual, there were two signals, and the following strategy was used for final classification. If both the signals were "N" type then final classification was "N". If any one signal was "T" type with probability greater than 0.5 then final classifications was "T". In case of a tie in probability value of "T" and "N", final classification was "T".

3. Results

The fractal dimension plot - slope of $\{log(L)$, - $log(N_{box}(L))\}$ - is presented in Fig.3 for signal-1 of Training data set nn05 [2]. The corresponding lacunarity plot with the best fit is presented in Fig.4.

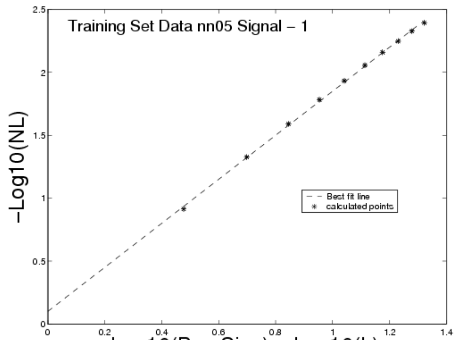

Figure 3. $\{log(L)$, - $log(N_{box}(L))\}$ plot along with the best fit line for signal-1 training data set nn05 [2]

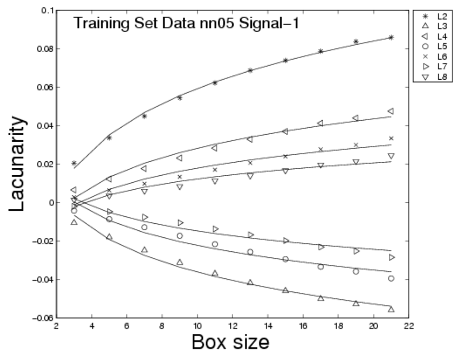

Figure 4. Lacunarity plots from $\Delta(L)_2$ to $\Delta(L)_8$ for signal-1 training data set nn05 [2]. The solid lines are best fit lines from equation-3

Table-1 presents the mean and standard deviation of the sixteen features and the independence between N and T groups in the training set. To establish the importance of different features, stepwise discriminant classification was performed on training set data by SPSS software. The most important feature selected as shown in Table-2, were a_8 and a_7. Figure5 presents the scatter plot for a_8 and a_7. Table-3 presents the leave-out-one cross validation result by using KNN classifier with all sixteen features (fractal dimension, mass dimension and fourteen lacunarity parameters - a_n and m_n - for $\Delta(L)_2$ to $\Delta(L)_8$) and with a_8 and a_7 only as features. The result of classification for both the sets of features was identical and the overall accuracy for signal1 and signal2 were 95 % and 85 % respectively. The result on the test set of challenge 2004 was 73.3 %. This result was obtained by using KNN classifier (k=3) and with sixteen fractal features. The classification result on test set with two most significant features, a_8 and a_7 using KNN classifier (k=3) was 66.6 %. As the exact classification for the test set was not known so the sensitivity, specificity etc., could not be calculated.

4. Discussion and conclusions

True Box counting algorithm has been used to estimate different fractal features. The plot of lacunarity matches with the concept of coefficient of variation for a statistically self similar set. The ECG signal is a statistically self afine fractal set and can be represented by y = f(x).

The reconstructed plain profile was self similar under different reduction ratios for x and y coordinate. This afine property can be taken care of, if we reconstruct the plain profile from the curve y = f(x). The fractal dimension plot presented in Fig.3 also showed an excellent fit on the data points. The data set nn05 from the training set was chosen arbitrarily.

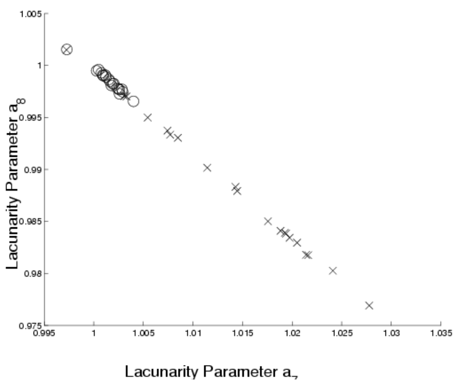

Figure 5. scatter plot of a_7 and a_8 (estimated by equation-3) from Training set. 'x' and 'o' are for 'N' type and 'T' type signals respectively. There are 20 'N' type and 20 'T' values.

It is known that the two sets with similar fractal dimensions may have different lacunarities. Thus lacunarities are more important as far as classification between fractal sets is concerned. The discriminant analysis also showed that lacunarity features a_8 and a_7 were most important as far as classification of ECG signals is concerned. Moreover, it was observed from the

Table 1. Independence between 'N' type and 'T' type - training set data. Fea.: Feature, G:Group, M:Mean, SD:Standard deviation, P:Probability,FD:Fractal Dimension,MD: Mass Dimension.

		Signal 1			Signal 2		
Fea.	G	M	SD	P $\leq$	M	SD	P $\leq$
FD	N	1.75	0.13		1.73	0.11	
	T	1.90	0.03	0.006	1.92	0.03	0.0005
MD	N	1.93	0.045		1.93	0.039	
	T	1.97	0.009	0.03	1.97	0.007	0.004
m_2	N	0.039	0.027		0.045	0.024	
	T	0.012	0.004	0.01	0.01	0.005	0.001
a_2	N	0.98	0.034		0.97	0.02	
	T	1.002	0.02	0.073	0.99	0.007	0.001
m_3	N	-0.02	0.011		-0.02	0.01	
	T	-0.006	0.003	0.003	-0.01	0.003	0.0005
a_3	N	1.012	0.014		1.02	0.009	
	T	0.99	0.009	0.025	1.002	0.004	0.0005
m_4	N	0.02	0.013		0.022	0.011	
	T	0.005	0.002	0.004	0.004	0.002	0.0005
a_4	N	0.98	0.015		0.978	0.012	
	T	0.99	0.005	0.006	0.998	0.003	0.0005
m_5	N	-0.016	0.009		-0.02	0.009	
	T	-0.004	0.002	0.003	0.001	0.001	0.0005
a_5	N	1.016	0.012		1.019	0.01	
	T	1.001	0.003	0.003	1.002	0.002	0.0005
m_6	N	0.013	0.008		0.015	0.007	
	T	0.003	0.001	0.003	0.003	0.001	0.0005
a_6	N	0.99	0.01		0.98	0.008	
	T	0.99	0.002	0.003	0.998	0.001	0.0005
m_7	N	-0.011	0.007		-0.01	0.006	
	T	-0.003	0.001	0.003	0.000	0.000	0.0005
a_7	N	1.013	0.009		1.015	0.008	
	T	1.001	0.002	0.002	1.002	0.001	0.0005
m_8	N	0.009	0.006		0.011	0.005	
	T	0.002	0.001	0.003	0.002	0.000	0.0005
a_8	N	0.99	0.007		0.99	0.006	
	T	0.99	0.001	0.002	0.99	0.000	0.0005

Table 2. Stepwise Discriminant analysis of fractal features on training set data. Lacunarity parameters a_8 and a_7 emerged as the most important features

Feature	F	Probability P $<$
a_8	47	0.00005
a_7	28	0.00005

Table 3. Classification summary for the training set using k-Nearest neighborhood Method. N:N-type, T:T-type, Tr.:True, Fa.:False, s1:Signal1, s2:Signal2, Sn:Sensitivity, Spc:Specificity, PV:Positive Predictive value, NV: negative predictive value, A:Accuracy

	Tr. N	Tr. T	Fa. N	Fa. T	Sn %	Spc %	PV %	NV %	A %
With all sixteen features									
s1	9	10	0	1	90	100	100	90.9	95
s2	8	9	1	2	80	90	88.8	81.8	85
With a_8 and a_7									
s1	9	10	0	1	90	100	100	90.9	95
s2	8	9	1	2	80	90	88.8	81.8	85

Fig.4 that the nonlinearity in the lacunarity plot decreases monotonically with higher mass moments. It is expected that this trend in decrease in nonlinearity will reach a saturation value as increasingly higher moments are used.

The leave-out-one cross validation result, presented in table-3 shows that classification error on T-type signal (i.e., abnormal) is less than N-type (i.e., normal) signals. This is significant in view of medical diagnosis. The cross validation result of KNN classifier with a_8 and a_7 and all sixteen features showed no difference on training set. This showed the non-significant contribution of other features on training set classification by KNN classifier. The overall accuracy achieved for signal1 and signal2 in the training set data were 95% and 85% respectively. Figure5 also showed a good separation between N-type and T-type signals with very little scatter in the training set data for the features a_8 and a_7.

However, the test set resulted in only 73.3 %. This test result is the outcome of final classification strategy adopted with a bias towards T-type signal and hence is conservative. As far as classification of signals is concerned, the fractal features based on lacunarities yielded good results as is evident from table-3. Moreover, the data in this work was not scaled. KNN classifier with Euclidean distance did not yield best classification with un-scaled data and this also deteriorated the classification accuracy in the result.

This preliminary investigation demonstrated that fractal features estimated by True Box Counting algorithm can result in highly significant classification between normal and abnormal ECG signals. Further lacunarities based on higher mass moments are more important than the other fractal features.

It is shown that the use of lacunarities based on higher mass moments ultimately reduced the feature space dimension for classification. Although this study considered a small set of ECG signals, it did show that fractal features yielded valuable classification information for ECG signals. Further investigation is necessary. A larger prospective study is planned.

Acknowledgements

The authors gratefully acknowledge the active support provided by the faculty of engineering, Al Tahadi University, Sirte, Libya in carrying out this research work. The authors would also like to thank George B. Moody, Harvard-MIT Division of Health Sciences and Technology MIT, USA for the encouragement through email and Ivanov Plamen, CAS Ctr for Polymer Studies, Boston University, USA for giving us the reprints of his research publication, which were timely and helpful.

References

[1] Blumenfeld R, Mandelbrot BB. Levy dusts, mittag-leffler statistics, mass fractal lacunarity, and perceived dimension. Physical Review E 1997;56:112–118.

[2] Physionet/computers in cardiology challenges February 2004;Downloaded from internet, http://www.physionet.org/challenge/.

[3] Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation 2000;101:e215–e220. Circulation Electronic Pages: http://circ.ahajournals.org/cgi/content/full/101/23/e215.

[4] Voss RF. Random fractals: Characterization and measurement. In Pynn R, Skjeltorp A (eds.), Scaling Phenomena in disordered systems.Plenum, New York,1986;.

[5] Ripley VB. k-nearest neighbour classification in statistical package R, package class 2004;URL `http://CRAN.R-project.org`.

Address for correspondence:

Dr. A. N. Esgiar
University of Al-Tahadi,
Faculty of Engineering,
PO Box 633,
Sirte - Libya.
anasser@autosyscntr.com