

Identification of Atrial Fibrillation Using Complex Network Similarity

Yajuan Zhang^{1,2}, Yuanyuan Wang^{1,2}

¹National Key Laboratory of ASIC and System, Fudan University, Shanghai, China

²Department of Electronic Engineering, Fudan University, Shanghai, China

Email: 11210720090@fudan.edu.cn, yywang@fudan.edu.cn

Received September 2012

ABSTRACT

We investigate the use of complex network similarity for the identification of atrial fibrillation. The similarity of the network is estimated via the joint recurrence plot and Hamming distance. Firstly, we transform multi-electrodes epicardium signals recorded from dogs into the recurrence complex network. Then, we extract features representing its similarity. Finally, epicardium signals are classified utilizing the classification and regression tree with extracted features. The method is validated using 1000 samples including 500 atrial fibrillation cases and 500 normal sinus ones. The sensitivity, specificity and accuracy of the identification are 98.2%, 98.8% and 98.5% respectively. This experiment indicates that our approach may lay a foundation for the prediction of the onset of atrial fibrillation.

Keywords: Recurrence Complex Network; Phase Space Reconstruction; Classification and Regression Tree; Atrial Fibrillation

1. Introduction

Atrial fibrillation (AF) is a common arrhythmia phenomenon in clinic [1,2]. During AF, chaotic electrical activities take place in the heart which result in an irregular heart rhythm. AF deteriorates the function of the whole heart and may cause embolic events and stroke which will induce high rates of morbidity and mortality. So it is important to detect AF and prevent it from becoming more serious.

With the development of nonlinear dynamics, the complex network [3-8] has become a popular method to analyze the dynamical properties of complex systems in the real world. Because the heart has been proven to a complex system, the heart electrical activity can be analyzed by using complex networks.

In [9], the distribution entropy of the point density in the two-dimensional reconstructed phase space was calculated as the feature of electrocardiogram to detect cardiac arrhythmia and the sensitivity, specificity and accuracy of the detection were 86.0%, 96.0% and 92.7% respectively. Because of the low sensitivity, a novel method needs to be proposed.

As we have known, for complex networks of chaotic electrical activities, the similarity between two networks will become weaker, so features representing the similarity between two recurrence complex networks can be calculated to identify AF. This will be proven with the

experiment.

2. Methods

2.1. Recurrence Complex Network

About how to construct complex networks from time series there have been already some approaches proposed. For example, Zhang and Small [3] introduced an algorithm to transform pseudo periodic time series into the complex network with each cycle of pseudo periodic series represented as a basic node. Lacasa *et al.* [4] proposed the visibility graph to map fractal time series into a network based on “visibility”. Yang and Yang [5] proposed to build the network from the correlation matrix of the time series. Though there are no embedding steps in the network construction algorithms mentioned above, it may become difficult to get necessary information from a high dimensional system. To remedy this defect, the recurrence complex network was proposed.

The recurrence network [6-9] constructs complex networks from a fractal time series. Each vector point of the reconstructed phase space is represented by a single node and the edge between two nodes is determined by the phase space distance. The adjacency matrix of the recurrence network is interpreted from the recurrence plot.

The recurrence plot starts from the phase space reconstruction which can be described as below. Take a fractal

time series $x(t) = \{x(t_1), x(t_2) \cdots x(t_N)\}$ as an example, where N is the length of the time series. If the embedding delay time is selected as τ and the embedding dimension selected as m , the vector point in the phase space can be represented as:

$\mathbf{X}(t) = \{x(t + \tau), x(t + 2\tau) \cdots x(t + (m-1)\tau)\}$. Then the phase space distance between vector points $\mathbf{X}(i)$ and $\mathbf{X}(j)$ is defined as:

$$d(\mathbf{X}(i), \mathbf{X}(j)) = \sum_{i,j=1}^m \|\mathbf{X}(i) - \mathbf{X}(j)\|. \quad (1)$$

Thus the recurrence plot of the time series $x(t)$ will be represented as:

$$\mathbf{R}_{i,j}(x(t)) = \begin{cases} 1, & \text{if } d(\mathbf{X}(i), \mathbf{X}(j)) < r, \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

where r is selected as the standard deviation of $d(\mathbf{X}(i), \mathbf{X}(j))$. Then the adjacent matrix \mathbf{A} will be described as:

$$\mathbf{A}_{i,j} = \mathbf{R}_{i,j}(x(t)) - \delta_{i,j}, \quad (3)$$

where δ is the identity matrix.

2.2. Feature Extraction

The features that represent the similarity between two complex networks are calculated based on the joint recurrence plot and Hamming distance. Firstly, we divide the multi-dimensional time series into shorter segments and reconstruct their phase space. Then, recurrence complex networks are constructed using Equations (2) and (3). The joint recurrence plot [10,11] is applied to test the similarity of two networks, which is defined as:

$$\mathbf{JR}_{i,j}(x^k(t), x^l(t)) = \mathbf{A}_{i,j}(x^k(t)) \cdot \mathbf{A}_{i,j}(x^l(t)), \quad (4)$$

where the time series $x^k(t)$, $x^l(t)$ are from the same divided segment, k and l correspond to the k th and l th time series.

If and only if the value is one at the node (i, j) in both recurrence complex networks, the value of the node (i, j) in the joint recurrence plot is one. $S = 0.5 * n(n-1)$ is defined as the size of the joint recurrence plot, where n is the number of time series in the segment, q_k and q_l are the recurrence ration of the recurrence plots $\mathbf{A}_{i,j}(x^k(t))$ and $\mathbf{A}_{i,j}(x^l(t))$ respectively. \mathbf{Z} [10] is defined as:

$$\mathbf{Z}_{k,l} = \frac{\sum_{i>j} \mathbf{JR}_{i,j}(x^k(t), x^l(t)) - Sq_k q_l}{\sqrt{Sq_k q_l (1 - q_k q_l)}}. \quad (5)$$

In order to analyze the similarity between two complex networks, \mathbf{Z} is normalized and shown in the form of pseudo-color map. As an example, pseudo-color maps of

\mathbf{Z} for AF and normal sinus are shown in **Figure 1**.

It is seen from **Figure 1** that there exists a great gap of \mathbf{Z} between nodes in diagonal and non diagonal rows. Thus a ration of the mean value of \mathbf{Z} is defined as:

$$RCN = \frac{\frac{1}{n} \sum_{k=1}^n \mathbf{Z}_{k,l}}{\frac{1}{n(n-1)} \sum_{k \neq l} \mathbf{Z}_{k,l}}. \quad (6)$$

The bigger RCN is, the weaker the similarity is.

Hamming distance is also used to test the similarity of two networks, it measures the edges that have to be changed to transform one network into the other. It is described as:

$$d(\mathbf{A}(k), \mathbf{A}(l)) = \sum \|\mathbf{A}_{i,j}(k) - \mathbf{A}_{i,j}(l)\|. \quad (7)$$

Based on Hamming distance, other three features are extracted to test the similarity of networks, which are the global mean Hamming distance (MH), the peak Hamming distance (PL) and the entropy of Hamming distance distribution ($ENTR$) respectively. Their definitions will be introduced in the below.

R is defined as the length of the reconstructed time series and Hamming distance for two constructed networks are calculated using Equation (7). Thus the definition of MH can be described as:

$$MH = \frac{1}{n * n} \sum_{k,l} d(\mathbf{A}(k), \mathbf{A}(l)) \frac{1}{R}. \quad (8)$$

MH responses the global mean Hamming distance of

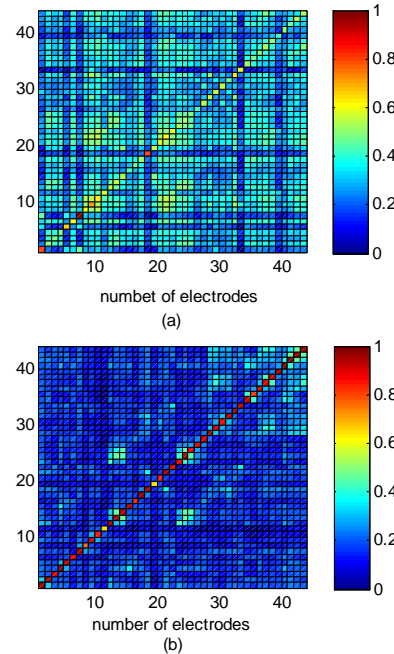


Figure 1. The pseudo-color map of \mathbf{Z} calculated from (a) AF segment and (b) normal sinus segment.

each segment, while L represents the local mean Hamming distance for each network constructed from the same segment. The probability distribution of L is calculated and rounded. It can be shown in bar plot. PL is defined as the distance which has the most probability:

$$PL = \arg \max_L p(L), \quad (9)$$

where $L_k = \frac{1}{n} \sum_k d(\mathbf{A}(k), \mathbf{A}(l)) \frac{1}{R}$ and $p(L)$ is the corresponding probability. For example, the distribution of L is shown in **Figure 2**.

The feature that measures the similarity is the entropy of Hamming distance distribution ($ENTR$). It is defined as:

$$ENTR = -\sum L^* \log(p(L)). \quad (10)$$

2.3. Classification and Regression Tree

The CART [12] approach is used in our research for the identification of AF with extracted features. The CART classification model utilizes tree-like structures to assign samples based on the set of feature quantifiers. It divide samples in parent nodes into descendent child nodes based on the if-then rule. The rule is obtained by training the sample set.

In the CART model, each parent node is optimally partitioned. The classification accuracy is measured by the decrease of the impurity [12] which is defined as:

$$\Delta im(s^*, tt) = im(tt) - p_L im(tt_L) - p_R im(tt_R), \quad (11)$$

where p_L denotes the proportion of cases in the left child node tt_L to those in the parent node tt while p_R denotes the proportion of cases in the right child node tt_R to those

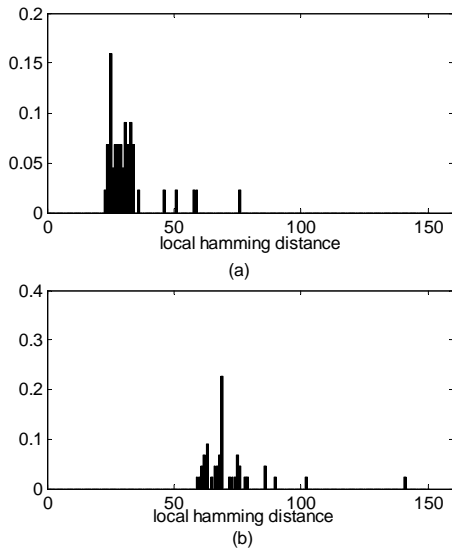


Figure 2. The distribution of L of (a) normal sinus segment and (b) AF segment.

in the parent node tt . im is the impurity function which is commonly selected as Gini diversity index criterion.

The best split s^* is selected to maximize $\Delta im(s^*, tt)$ and s^* is obtained by solving

$$\Delta im(s^*, tt) = \arg \max_s \Delta im(s, tt) - \arg \max_s (im(tt) - p_L im(tt_L) - p_R im(tt_R)) \quad (12)$$

3. Experiments and Results

The method is implanted in Mat lab R2011b and the platform is a computer workstation with a 3.07 GHz CPU and 32 G memory.

The data are from dog experiments, in which the perfusion of acetylcholine is used to induce AF in anesthetized open-chest dogs with the electrical stimulation of left atrium appendage. 128 unipolar electrodes are placed on the atrial epicardial surface and thus a recording of 128 channels is obtained.

Signals are filtered between 3.5 - 600 Hz, and the sampling frequency is 2 kHz with 16-bit resolution. Due to that the anterior right atrium signals can well represent atrial activities, 44 recordings from the right atrium surface, marked from 1 to 44, are used to construct the recurrence complex networks. Signals are divided into short segments which are used as samples in the classification experiment.

The classification and regression tree integrated with 10-fold cross-validation is used in our study. The classification performance parameters used to measure the performance of methods are sensitivity (SE), specificity (SP) and accuracy (AC). They are defined as:

$$SE = \frac{TP}{TP + FN}, \quad SP = \frac{TN}{TN + FP},$$

$$AC = \frac{TP + TN}{TP + FN + TN + FP}, \quad (13)$$

where TP is the number of AF cases being correctly recognized as AF, FN is the number of AF cases being wrongly recognized as normal sinus ones, TN is the number of normal sinus cases being truly recognized as normal sinus ones, and FP is the number of normal sinus cases being wrongly recognized as AF ones.

1000 samples are used in the classification experiment with 500 AF cases and 500 normal sinus ones and the network similarity parameters are calculated as classification features. The time interval of the segment may lead to the difference of the classification performance. Here we compare the classification results with various time intervals. **Table 1** shows the classification performance with the interval of 100, 150, 200 and 250 ms respectively. The time interval is limited to 250 ms because of the large computing amount.

It is shown in **Table 1** that the classification perfor-

Table 1. The classification performance with various intervals.

Interval (ms)	SE (%)	SP (%)	AC (%)
100	54.6	85.2	69.3
150	59.1	82.5	70.0
200	61.1	71.8	65.7
250	72.6	82.4	77.7

mance is not satisfied for all time intervals.

However, with the increase of the time interval, the performance becomes better (except *SP*). If we enlarge the time interval to get a better result, the computing amount will be much bigger. Therefore, in order to enlarge the time interval with the computing amount remaining unchanged, we have to down sample signals. The down sampling frequency is determined by the power spectral analysis. We calculate the power spectrum of signals and find that the frequency range of signals is about 0 - 200 Hz, so the down sampling frequency is determined as 400 Hz according to the Nyquist Criteria.

We compare the classification performance with the down sampling frequency of 400 Hz and the time interval of 500, 750, 1000 and 1250 ms respectively. The results are shown in **Figure 3**. It has shown that the optimal result is obtained with the interval of 750 ms. It illustrates that it's not the longer the time interval is, the better is the performance. The classification performance will become worse with the time interval longer than 750 ms.

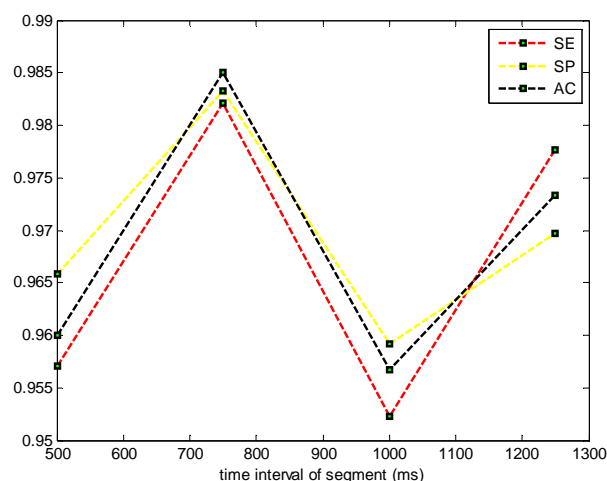
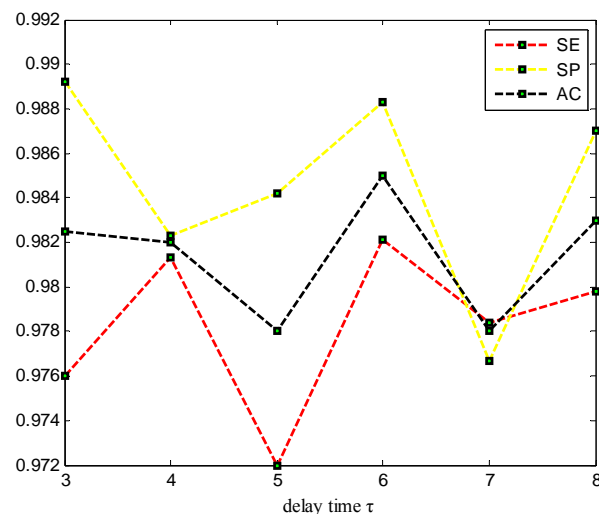
The classification reliability is also compared with various embedding delay time τ . The result is shown in **Figure 4**. It has been illustrated that the more reliable identification of AF is obtained with τ of 6.

Table 2 shows comparison results of our method with the method in [9]. The significant improvement in the sensitivity, specificity and accuracy has been seen in our method compared with the method in [9]. For example, our method has 12.2% improvement in the sensitivity.

4. Conclusions and Discussions

In our study, we utilize the recurrence complex network to extract similarity parameters between two recurrence complex networks to recognize AF segments from normal sinus cases. In the classification experiments with down sampling frequency of 400 Hz and various time intervals, the best classification performance is obtained with the time interval selected as 750 ms, and the sensitivity, specificity and accuracy are 98.2%, 98.8% and 98.5% respectively.

The improved classification performance means that our approach can detect AF more accurately and sensi-

**Figure 3. The classification performance with down sampling frequency of 400 Hz and various time intervals.****Figure 4. The classification performance with various delay time τ .****Table 2. The classification performance comparison.**

Method	SE (%)	SP (%)	AC (%)
Our method	98.2	98.8	98.5
Method in [9]	86.0	96.0	92.7

tively. However, the improvement of the performance increases the computing amount. For signals used in this study are 44 dimensions, the computing amount is much larger than that of method in [9].

Besides the computing amount, there are still two disadvantages in our method. One is that the method ignores the other cardiac arrhythmia phenomena, which will deteriorate the validity of our method. The other is that it is unknown whether or not the approach is suitable for human body, which will be our future studies.

5. Acknowledgements

This work is supported by the National Natural Science Foundation of China (Grant No. 61271071 and No. 11228411), the National Key Technology R&D Program of China (No. 2012BAI13B02) and Specialized Research Fund for the Doctoral Program of Higher Education of China (No.20110071110017).

REFERENCES

- [1] A. L. Waldo, "Mechanism of Atrial Fibrillation," *Journal of Cardiovascular Electrophysiology*, Vol. 14, No. 12, 2003, pp. s267-s274.
<http://dx.doi.org/10.1046/j.1540-8167.2003.90401.x>
- [2] S. Nattel, A. Shiroshita-Takeshita, B. Brundel and L. Rivard, "Mechanisms of Atrial Fibrillation: Lessons from Animal Models," *Progress in Cardiovascular Diseases*, Vol. 48, No. 1, 2005, pp. 9-28.
<http://dx.doi.org/10.1016/j.pcad.2005.06.002>
- [3] J. Zhang and M. Small, "Complex Network from Pseudoperiodic Time Series: Topology versus Dynamics," *Physical Review Letters*, Vol. 96, No. 23, 2006, Article ID: 238701.
<http://dx.doi.org/10.1103/PhysRevLett.96.238701>
- [4] L. Lacasa, B. Luque, F. Ballesteros, J. Luque and J. C. Nuno, "From Time Series to Complex Networks: The Visibility Graph," *PNAS*, Vol. 105, No. 13, 2008, pp. 4972-4975. <http://dx.doi.org/10.1073/pnas.0709247105>
- [5] Y. Yang and H. Yang, "Complex Network-Based Time Series Analysis," *Physica A*, Vol. 387, 2008, pp. 1381-1386. <http://dx.doi.org/10.1016/j.physa.2007.10.055>
- [6] Z. Gao and N. Jin, "Complex Network from Time Series Based on Phase Space Reconstruction," *Chaos*, Vol. 19, No. 3, 2009, Article ID: 033137.
<http://dx.doi.org/10.1063/1.3227736>
- [7] N. Marwan, J. F. Donges, Y. Zou, R. V. Donner and J. Kurths, "Complex Network Approach for Recurrence Analysis of Time Series," *Physics Letters A*, Vol. 373, No. 9, 2009, pp. 4246-4254.
<http://dx.doi.org/10.1016/j.physleta.2009.09.042>
- [8] L. Uldry, V. Jacquemet, N. Virag, L. Kappenberger and J. M. Vesin. "Estimating the Time Scale and Anatomical Location of Atrial Fibrillation Spontaneous Termination in a Biophysical Model," *Medical and Biological Engineering and Computing*, Vol. 50, 2012, pp. 155-163.
<http://dx.doi.org/10.1007/s11517-011-0859-3>
- [9] R. Sun, Y. Wang, S. Yang and Z. Fang. "Detecting Cardiac Arrhythmias Based on Phase Space Analysis," *Journal of Biomedical Engineering*, Vol. 25, No. 4, 2008, pp. 934-937.
- [10] Y. Hirata and K. Aihara, "Identifying Hidden Common Causes from Bivariate Time Series: A Method Using Recurrence Plots," *Physical Review E*, Vol. 81, 2010, Article ID: 016203.
<http://dx.doi.org/10.1103/PhysRevE.81.016203>
- [11] N. Marwan, M. C. Romano, M. Thiel and J. Kurths, "Recurrence Plots for the Analysis of Complex Systems," *Physics Reports*, Vol. 438, 2007, pp. 237-329.
<http://dx.doi.org/10.1016/j.physrep.2006.11.001>
- [12] H. Yanga, S. Bukkapatnamb, T. Leb and R. Komanduric, "Identification of Myocardial Infarction (MI) Using Spatio-Temporal Heart Dynamics," *Medical Engineering & Physics*, Vol. 34, 2012, pp. 485-497.
<http://dx.doi.org/10.1016/j.medengphy.2011.08.009>