Belief χ^2 Divergence-based Dynamical Complexity Analysis for Biological Systems

Lang Zhang and Fuyuan Xiao

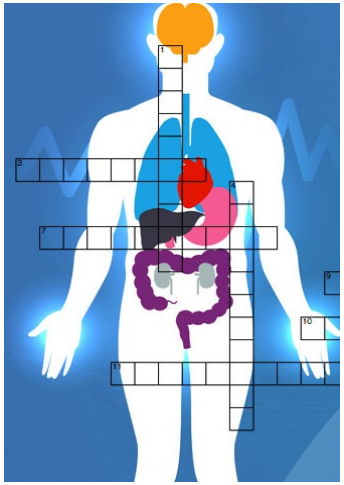
Chongqing University, Chongqing, China

4th International Conference on Machine Learning for Cyber Security

(ML4CS 2022)

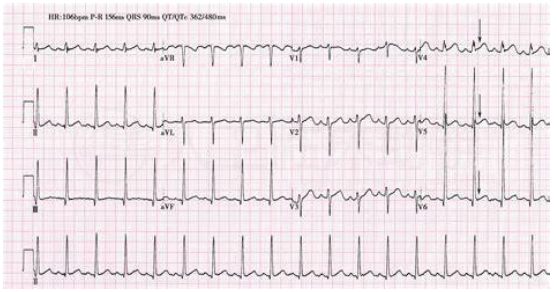
Outline

Background



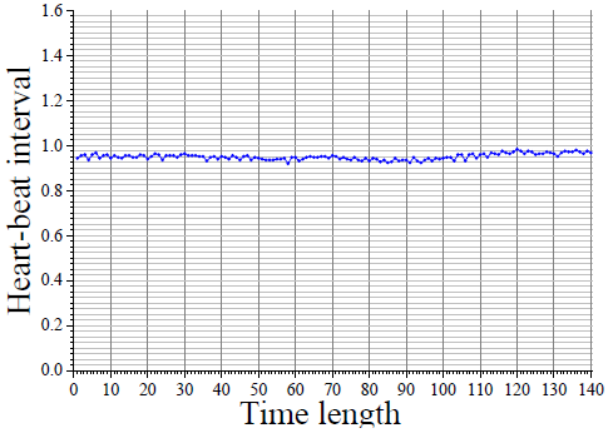
Biological systems

EEG signals



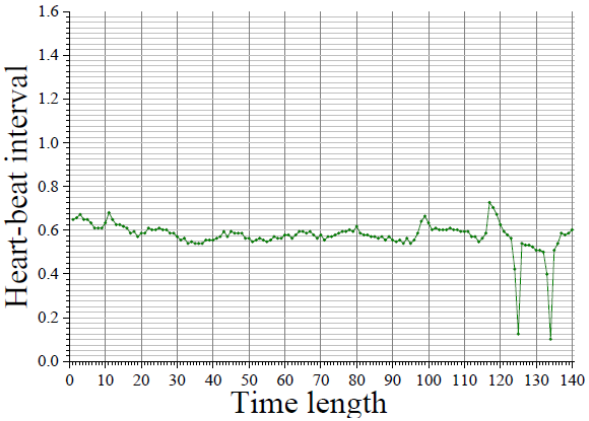
Cardiac inter-beat signals

Background



Pathological subjects

More information



Healthy subjects

- **Recent works for uncertain information management:**

Deng entropy

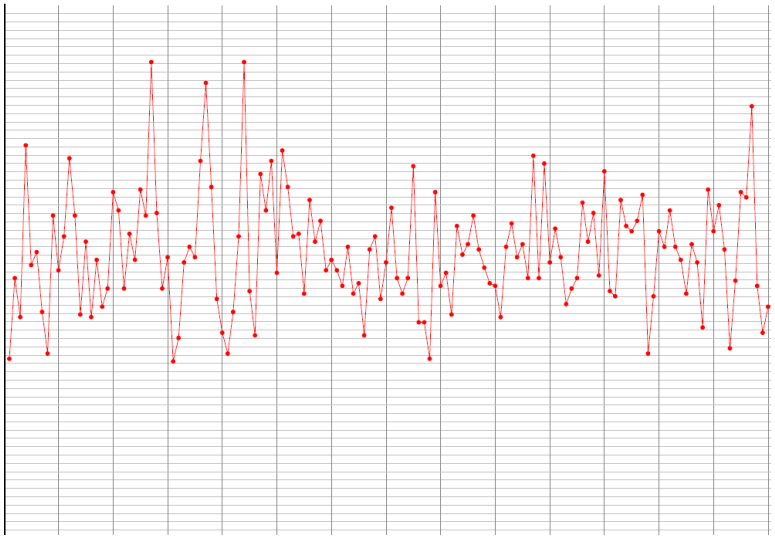
Z-network

Information quality

Belief entropy

TDBF

Motivation



Sophisticated structure

↓

Feature extraction based on Dempster-Shafer evidence theory — BPAs

↓

Process the time series data points whether they are on the boundaries of time slice

Preliminaries

- **Dempster-Shafer evidence theory**

Definition 1 (Framework of discernment).

Let Θ be a set that consists of r mutually exclusive and collectively exhaustive events,

$$\Theta = \{e_1, e_2, \dots, e_i, \dots, e_r\}$$

which indicates the framework of discernment. The power set 2^Θ is used to describe uncertainty which can be defined as follows:

$$2^\Theta = \{\emptyset, \{e_1\}, \dots, \{e_r\}, \{e_1, e_2\}, \dots, \{e_1, e_2, \dots, e_h\}, \dots, \Theta\},$$

where $\emptyset$ indicates the empty set.

Definition 2 (Mass function).

Based on the frame of discernment Θ, m as a mass function, also known as BPA, is a mapping from 2^Θ to [0,1] which is defined as:

$$m: 2^\Theta \to [0,1].$$

Because events must arise from propositions in the framework of discernment and empty set is not the cause of the events, it abides the rule of

$$\sum_{E \in 2^\Theta} m(E) = 1 \text{ and } m(\emptyset) = 0.$$

If $m(E) > 0$ E is a focal element.

Preliminaries

- ***SEB*χ^2 divergence measure**

Let m_1 and m_2 be two BPAs. The $SEB\chi^2$ divergence measure [30] can be defined as:

$$D_{SEB\chi^2}(m_1, m_2) = \frac{1}{2}\left[D_{EB\chi^2}\left(m_1, \frac{m_1 + m_2}{2}\right) + D_{EB\chi^2}\left(m_2, \frac{m_1 + m_2}{2}\right)\right]. \tag{5}$$

Symmetry

where

$$D_{EB\chi^2}(m_1, m_2) = \sqrt{\left(\frac{m_1(\theta)}{\sqrt{m_2(\theta)}} - \sqrt{m_2(\theta)}\right)' \Psi \left(\frac{m_1(\theta)}{\sqrt{m_2(\theta)}} - \sqrt{m_2(\theta)}\right)}, \tag{6}$$

with $\theta \in \Theta$, and

$$\Psi(F_i, F_j) = \frac{2^{|F_i \cap F_j|} - 1}{2^{|F_i|} - 1} \cdot \frac{2^{|F_i \cap F_j|} - 1}{2^{|F_j|} - 1}. \tag{7}$$

Consider the number of element

F_i and F_j represent m_1 and m_2 $(i, j = 1, 2, \ldots, 2^{n-1})$. $|\cdot|$ indicates the cardinality of a BPA. Ψ can be regarded as correlation coefficient [34].

The Proposed Method (DCA)

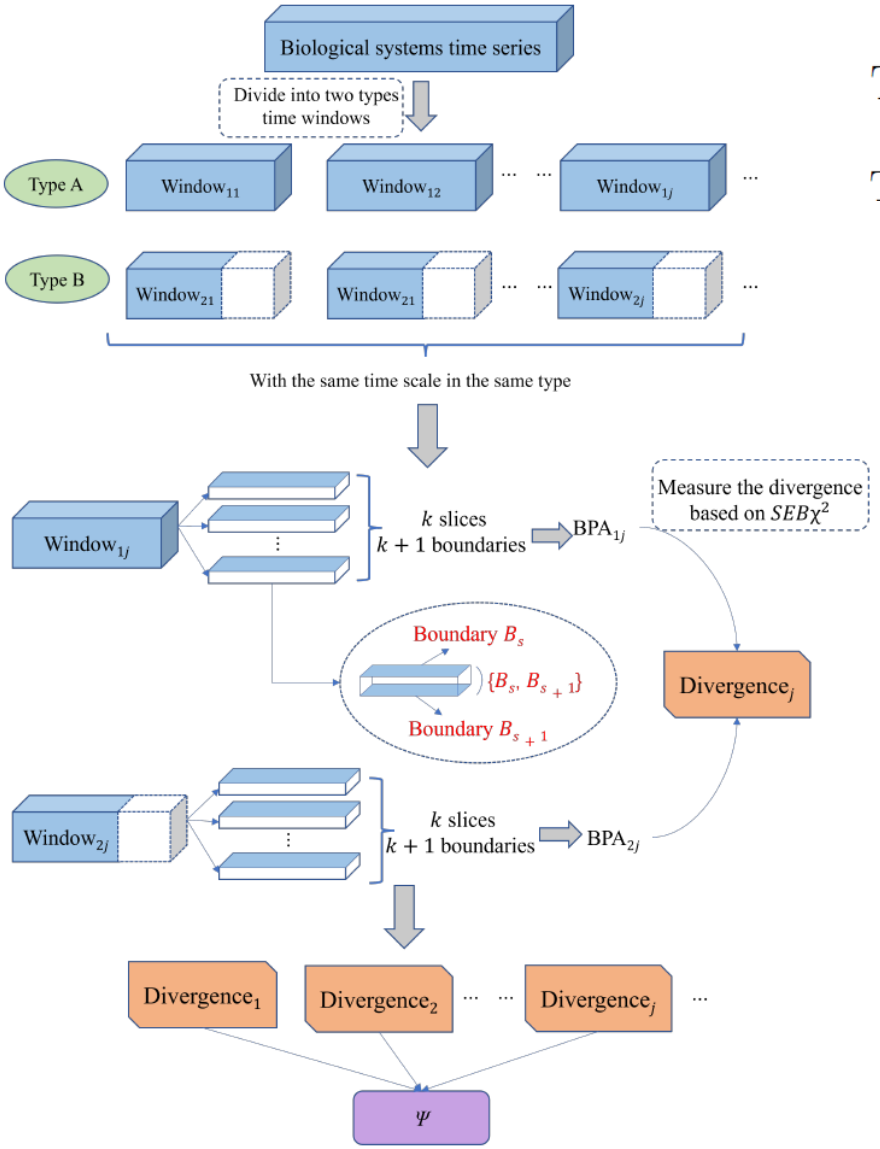


Fig. 1: Flowchart of the DCA algorithm for biological systems.

- **Two lists of consecutive non-overlapping time windows**

$$\text{Type A } w_{Aj}^{(\tau)} = \{x_{(j-1)\tau+1}, \cdots, x_{(j-1)\tau+\tau}\}$$

$$\text{Type B } w_{Bj}^{(\tau)} = \{x_{(j-1)\tau+1}, \cdots, x_{(j-1)\tau+\upsilon}\}$$

- **Focal element of BPA**

$$m_{ij}(\{B_s, B_{s+1}\}) = \frac{p}{|w_{ij}|}, \qquad i \in \{A, B\},$$

$$m_{ij}(\{B_s\}) = \frac{q}{|w_{ij}|} \qquad i \in \{A, B\}.$$

- ***Div$_j$* in each corresponding window**

$$Div_j = D_{SEB\chi^2}(m_{Aj}, m_{Bj}).$$

- **The average divergence represents the complexity of a biological system time series Ψ**

$$\Psi = \frac{\sum_{i=1}^{N/\tau} Div_i}{N/\tau}.$$

The Proposed Method

The pseudocode of dynamical complexity analysis algorithm for biological systems based on $SEB\chi^2$ divergence is shown in Algorithm 1.

Algorithm 1: Complexity analysis algorithm for biological systems based on $SEB\chi^2$ divergence

Input: Biological systems time series $\{x_i\} = \{x_1, \ldots, x_N\}$;
Output: Complexity result Ψ

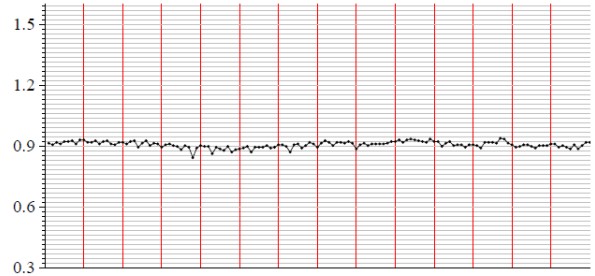
Split the time series $\{x_i\}$ into two types of windows $\left\{w_{Aj}^{(\tau)}\right\}$ and $\left\{w_{Bj}^{(\tau)}\right\}$;
Determine the lower and upper sides of time interval $\{x_{min}, x_{max}\}$;
Divided each time window into k slices;
Count the number of data points on or between boundaries;
for $i=1; i \leq N/\tau$ **do**
    Figure out the BPAs m_{1i} and m_{2i} of each time window by using the Eq. (8) and Eq. (9);
end
for $i=1; i \leq N/\tau$ **do**
    Calculate the divergence Div_i in each corresponding window by using Eq. (10);
end
Calculate the complexity of biological systems time series Ψ by using Eq. (11);
return Ψ.

Feature extraction

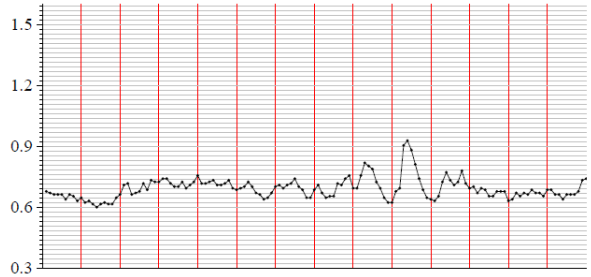
Complexity measurement

Simulation Experiments

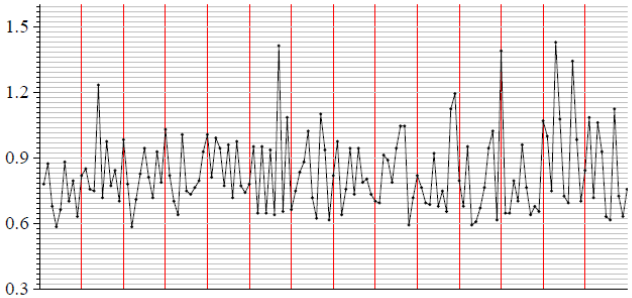
Cardiac inter-beat time series



Original time series of CHF



Original time series of Healthy



Original time series of AF

In this study, cardiac inter-beat interval time series is applied to demonstrate the feasibility of DCA algorithm for biological systems complexity analysis. The data is selected from the databases on PhysioNet as follows:

- BIDMC Congestive Heart Failure Database (CHF);
- MIT-BIH Normal Sinus Rhythm Database (Healthy);
- Long Term AF Database (AF).

Simulation Experiments

Data processing

- Each subject is truncated into 5 sets interbeat interval time series by utilizing the first 500 data points from 10,000 data points.
- Hence, there are 240 sets inter-beat interval time series. Specifically, 75, 90 and 75 records are from CHF Healthy and AF, respectively.
- The data points $\{x_i\}$ are ranked and split into 1000 segments. To release the influence of noise and detection error, the $1st$ and $999th$ 1000-quantiles of the ranked segments are regarded as x_{min}= 0.3 and x_{max} = 1.6

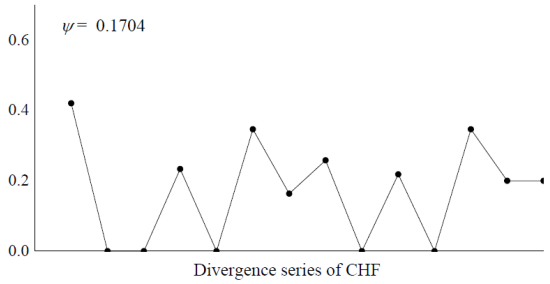
Simulation Experiments

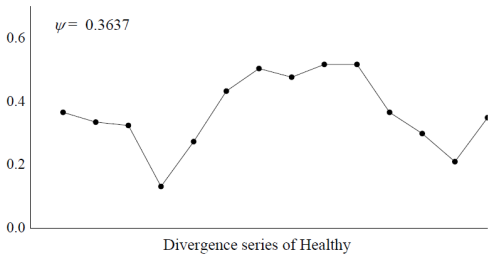
- Table 1 shows the 14 divergence values for each time window of data sets, respectively.

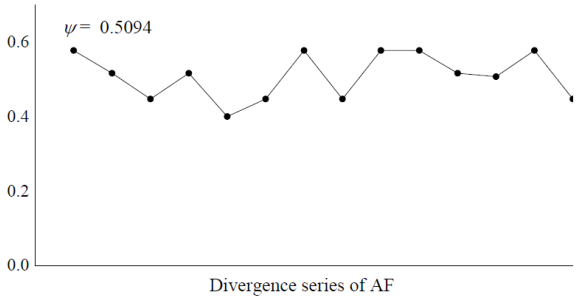
Table 1: The divergence value of CHF, Healthy and AF in each time window.

Subject	Win_1	Win_2	Win_3	Win_4	Win_5	Win_6	Win_7
CHF	0.4201	0.0000	0.0000	0.2335	0.0000	0.3464	0.1633
Healthy	0.3651	0.3342	0.3237	0.1309	0.2725	0.4320	0.5033
AF	0.5773	0.5164	0.4472	0.5164	0.4000	0.4472	0.5773
Subject	Win_8	Win_9	Win_{10}	Win_{11}	Win_{12}	Win_{13}	Win_{14}
CHF	0.2582	0.0000	0.2182	0.0000	0.3464	0.2000	0.2000
Healthy	0.4761	0.5164	0.5164	0.3651	0.2981	0.2093	0.4381
AF	0.4472	0.5773	0.5773	0.5164	0.5071	0.5773	0.4472

- Fig. 2 shows the three original time series and divergence series, respectively.







Conclusion

Contribution:

- Biological systems time series data is converted into mass function by using the D-S evidence theory, where feature of data can be extracted.
- The proposed DCA algorithm proposes an effective way to figure out the complexity of time series data in biological systems by generating BPAs and measure the average divergence of them.
- An application for pathological states analysis in cardiac inter-beat interval time series is carried out to illustrate the effectiveness of DCA algorithm.

Future work:

- The time complexity of the DCA algorithm for biological systems should be addressed to adapt to real-time data flexibly.

THANK YOU VERY MUCH!

Any questions and comments are welcome!

Email address: zhanglang_cqu@163.com