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

Lang Zhang and Fuyuan Xiao

School of Big Data and Software Engineering, Chongqing University, Chongqing 401331, China
xiaofuyuan@cqu.edu.cn; doctorxiaofy@hotmail.com

Abstract. Physiological signals contain the information of physical state in healthy systems. Especially, the complexity of dynamical time series data is a valid indicator to measure the pathological states. However, how to quantify the complexity of physical signals is still an open issue. In this paper, a novel complexity analysis algorithm based on divergence, called DCA, is proposed to figure out the complexity of biological system time series. Specifically, DCA algorithm splits biological systems time series data into different slices with boundaries. In addition, the feature of each time series data is extracted by converted into the basic probability assignments (BPAs) based on the Dempster-Shafer (D-S) evidence theory. DCA algorithm considers that the average divergence of BPAs indicates the complexity in a piece of time series. Moreover, an application in cardiac inter-beat interval time series is carried out to demonstrate the effectiveness of the proposed algorithm, which performs well in a pathological states analysis issue.

Keywords: D-S evidence theory · Belief χ^2 divergence · Dynamical complexity analysis · Biological system.

1 Introduction

Biological systems produce valuable information with time series data, which contributes to the field of pathological researches. Specifically, complexity analysis of biological system time series reflects the changing environment of the patients physical state [1–3]. The complexity of the biological systems data can be measured effectively with entropy of information [4, 5]. In time series, the more information there is, the larger complexity it should be. Moreover, it means that the information is more uncertain. So, the complexity of a piece of time series can be measured by figuring out the uncertainty of the information. Besides, uncertainty measure is widely applied in multi-source information fusion for decision

* supported by the National Natural Science Foundation of China (No. 62003280), Chongqing Talents: Exceptional Young Talents Project (No. cstc2022ycjh-bgzxm0070), and Chongqing Overseas Scholars Innovation Program (No. cx2022024).
Corresponding author: Fuyuan Xiao (xiaofuyuan@cqu.edu.cn;doctorxiaofy@hotmail.com)

making [6], fault analysis [7], and so on [8]. Nevertheless, it is still a challenge to measure the uncertainty of information which may be conflicting [9]. Hence, several well-known works has been proposed to manage multi-source information, including Deng entropy [10], belief entropy [11], Z-network [12], TDBF [13], information quality [14], and so on [15–17].

In this study, it is considered that D-S evidence theory [18–20] addresses a uncertainty problem in a flexible way [21–23]. In this case, the D-S evidence theory is taken into account to process the biological systems time series data by means of presenting them in the format of BPAs. BPAs are able to illustrate the uncertainty feature of data points by assigning probabilities to different categories [24, 25]. In order to measure the discrepancy of BPAs, following aspects are investigated by scholars, including distance [26], correlation coefficient and divergence. Particularly, Xiao et al. [27] proposed a uniform BJS divergence-based method to consider both subjective weights and objective weights. Zhu et al. [28] considered the number of possible hypotheses with Rényi divergence. Wang et al. proposed IVIFJS divergence [29] to take the weight of membership and non-membership into account. Recently, Zhang and Xiao [30] proposed $SEB\chi^2$ divergence to measure the difference between belief function by focusing the discrepancy and correlationship of both singleton sets and multi-element sets. As for time series of biological systems, they often have more sophisticated structure. Hence, there is a limitation that it is difficult to measure the complexity of time series effectively. In this case, $SEB\chi^2$ is taken into account to be applied to process the time series data points whether they are on the boundaries of time slices. At the same time, the inner feature of biological systems can be illustrated.

In this work, a novel complexity analysis algorithm based on $SEB\chi^2$ divergence, called DCA, is proposed to measure the complexity of time series in biological systems. DCA algorithm divides biological systems data into multiple BPAs with different time scales. Then, the complexity of time series is obtained by averaging the divergence of corresponding BPAs. Furthermore, a specific application in cardiac inter-beat interval time series is carried out to demonstrate the effective performance in analyzing the complexity in a real issue.

Main contributions are presented as follows:

• 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.

The structure of this work shows as follows: In Section 2, the basic concepts of D-S evidence theory and divergence are introduced. A novel DCA algorithm in complexity analysis for biological systems is derived in Section 3. Section 4 puts forward an implement to illustrate the effective performance of the proposed algorithm in a real issue. Section 5 makes a conclusion of this paper.

2 D-S Evidence Theory and Divergence Measure

Some fundamental concepts are briefly presented in this section, including D-S evidence theory [18, 19] and $SEB\chi^2$ divergence measure [30].

Definition 1 (Framework of discernment).

Let the discernment Θ be a finite set which can be defined as:

$$\Theta = \{f_1, f_2, \ldots, f_n\}. \tag{1}$$

Then, its power set 2^Θ can be defined as [31]:

$$2^\Theta = \{\varnothing, \{f_1\}, \ldots, \{f_n\}, \{f_1, f_2\}, \ldots, \{f_1, f_2, \ldots, f_h\}, \ldots, \Theta\}, \tag{2}$$

where $\varnothing$ indicates the empty set.

Definition 2 (Mass function).

Based on discernment Θ, the mass function m [32, 33] can be defined as:

$$m : 2^\Theta \to [0, 1], \tag{3}$$

with the rule of

$$\sum_{E \in 2^\Theta} m(E) = 1 \quad \text{and} \quad m(\varnothing) = 0. \tag{4}$$

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

Definition 3 ($SEB\chi^2$ divergence measure).

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

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

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

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

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].

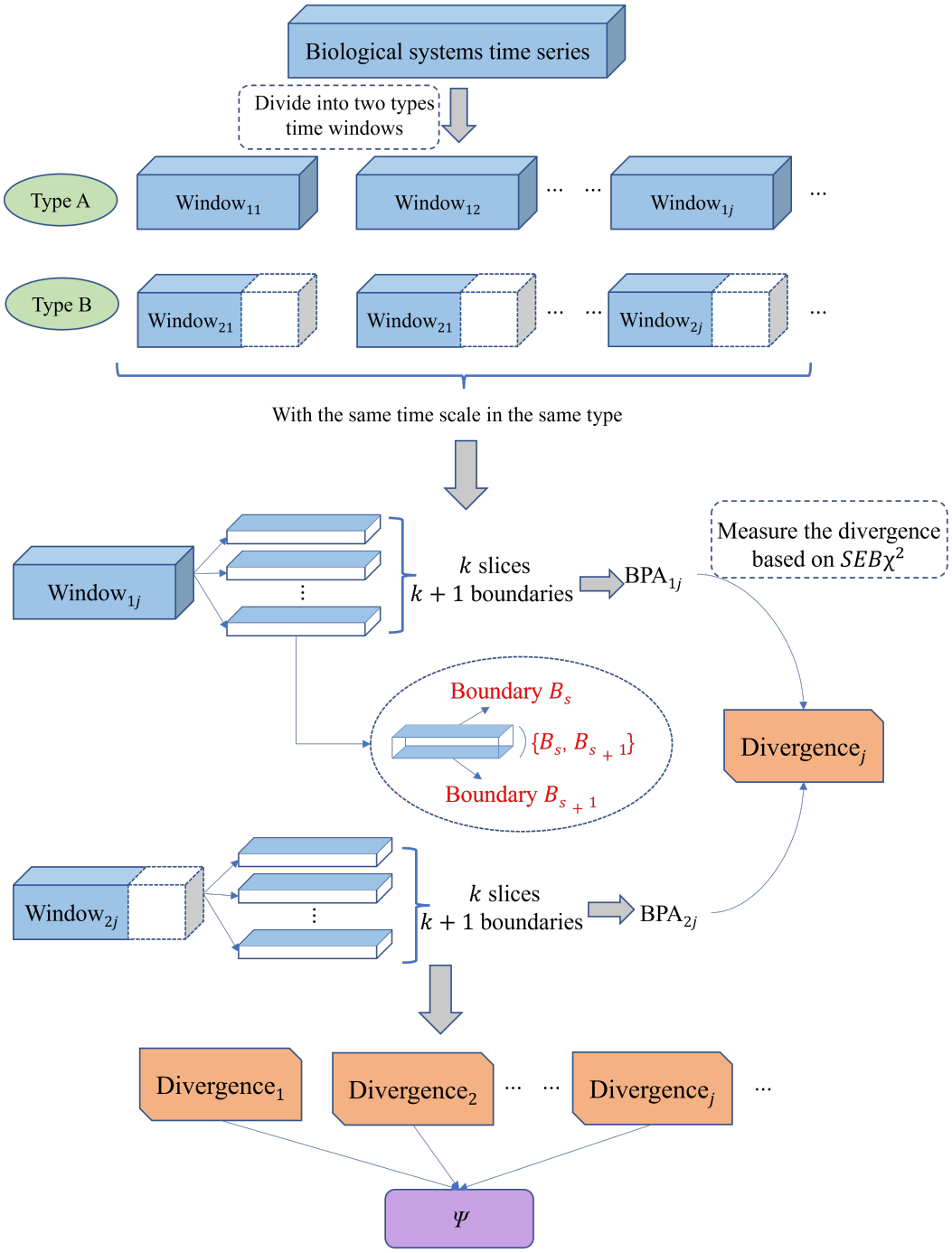

Fig. 1: Flowchart of the DCA algorithm for biological systems.

3 $SEB\chi^2$-based Dynamical Complexity Analysis Algorithm for Biological Systems

The dynamical complexity analysis algorithm based on $SEB\chi^2$ divergence is introduced in this section which consists two steps. The flowchart of DCA algorithm for biological systems is shown in Fig. 1.

In the first step, biological systems time series $\{x_i\} = \{x_1, \ldots, x_N\}$ with length N is divided into two lists of consecutive non-overlapping time windows as type A $\left\{w_{Aj}^{(\tau)}\right\}$ and Type B $\left\{w_{Bj}^{(\tau)}\right\}$. Type A $w_{Aj}^{(\tau)} = \left\{x_{(j-1)\tau+1}, \ldots, x_{(j-1)\tau+\tau}\right\}$ is of length τ. Besides, j is the window index which ranges from 1 to N/τ. It can be regarded that type B is the truncation of type A in each window as $w_{Bj}^{(\tau)} = \left\{x_{(j-1)\tau+1}, \ldots, x_{(j-1)\tau+\upsilon}\right\}$, where $\upsilon < \tau$.

The time interval of each window ranges from x_{min} to x_{max} where x_{min} and x_{min} indicate the lower and the upper boundaries, respectively, of time series $\{x_i\}$. The time interval is equally split into k slices. Each slice contains two boundaries as B_s and B_{s+1}, which represents the specific state. If data points are in the same slice, then it can be considered that they are in the same state.

Let the total number of x_i over w_{Aj} or w_{Bj} between B_s and B_{s+1} be p. Then, the focal element of BPA based on each time window can be defined as:

$$m_{ij}(\{B_s, B_{s+1}\}) = \frac{p}{|w_{ij}|}, \qquad i \in \{A, B\}, \tag{8}$$

where indicates the length of time series. In addition, if data points of length q fall on the border B_s coincidentally, the focal element can be defined as follows:

$$m_{ij}(\{B_s\}) = \frac{q}{|w_{ij}|} \qquad i \in \{A, B\}. \tag{9}$$

In the second step, the divergence Div_j in each corresponding window is figured out based on $SEB\chi^2$ divergence measure:

$$Div_j = D_{SEB\chi^2}(m_{Aj}, m_{Bj}). \tag{10}$$

Finally, the average divergence represents the complexity of a biological system time series Ψ:

$$\Psi = \frac{\sum_{i=1}^{N/\tau} Div_i}{N/\tau}. \tag{11}$$

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

4 Application

In this section, the biological systems time series data is described, which shows the way of selecting valid data points. Next, an implement of DCA algorithm for biological systems is carried out to shows the effective performance in specific time series.

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.

4.1 Data Description

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 [35] as follows:

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

All the above databases are long-term ECG (Electrocardiography) databases with 20-24 hours record. Specifically, the numbers of subject are 15, 18 and 84. As for CHF and Healthy databases, each subject is truncated into 5 sets inter-beat interval time series by utilizing the first 500 data points from 10,000 data points. As for AF data base, 75 records are adopted according to the annotation in PhysioNet. Because the lengths of them exceed 500 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.

Next, the time series is processed. First, the data points $\{x_i\}$ are ranked and split into 1000 segments. To release the influence of noise and detection error, the 1_{st} and 999_{th} 1000-quantiles of the ranked segments are regarded as $x_{min} = 0.3$ and $x_{max} = 1.6$.

4.2 Implement of DCA Algorithm for Biological Systems

Three specific instances are carried out to demonstrate the process of DCA algorithm for biological systems. Three representative biological systems time

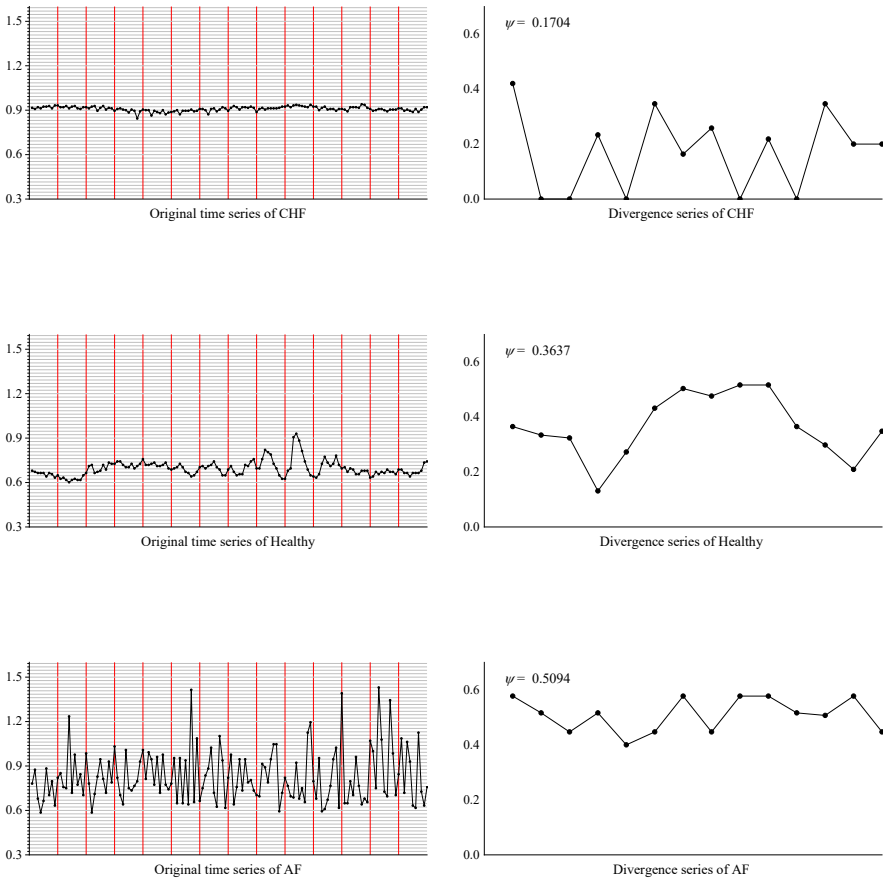

Fig. 2: Demonstration of DCA algorithm for biological systems on specific instances.

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

series from CHF, Healthy and AF are taken into consideration. To simplify the experiment, each time series is analyzed with 140 data points, whose parameters are at $\tau = 10$, $\upsilon = 5$ and $k = 55$. In this case, each time series will be split into 14 time windows.

Table 1 shows the 14 divergence values for each time window of data sets, respectively. Fig. 2 shows the three original time series and divergence series, respectively. The resulting complexity values Ψ of the three sets above are obtained as 0.1704 of CHF, 0.3637 of Healthy and 0.5094 of AF.

From the original time series of three data sets, it shows that the CHF subject has the lowest fluctuation while AF subject has the highest fluctuation in cardiac inter-beat interval time series. As divergence illustrates the inner difference of a single time window, the divergence of three sets should follows: CHF subject $<$ Healthy subject $<$ AF subject. In this case, divergence can be used as an indicator of complexity.

From the information above, DCA algorithm can effectively analyze the complexity of biological systems time series data of different types.

5 Conclusion

As biological systems usually produces time series, the complexity analysis on time series are significant. This research shed new light on the dynamical complexity analysis based on $SEB\chi^2$ divergence for biological systems. The main innovation point of the proposed method was that it took inner divergence of a time series into consideration. Also, the data points on boundaries and in slices were dealt differently by generating BPAs. In addition, the effectiveness of DCA algorithm for biological systems was demonstrated by applying it in cardiac inter-beat interval time series. Moreover, DCA algorithm provided a novel way to address the biological systems problems in physical state analysis. In summary, DCA algorithm had considerable abilities in dynamical complexity analysis for biological systems. In the future study, the time complexity of the DCA algorithm for biological systems should be addressed to adapt to real-time data flexibly.

References

1. H. Cui, L. Zhou, Y. Li, B. Kang, Belief entropy-of-entropy and its application in the cardiac interbeat interval time series analysis, Chaos, Solitons & Fractals 155 (2022) 111736.
2. C. F. Hsu, S.-Y. Wei, H.-P. Huang, L. Hsu, S. Chi, C.-K. Peng, Entropy of entropy: Measurement of dynamical complexity for biological systems, Entropy 19 (10) (2017).
3. Z. Cao, W. Ding, Y.-K. Wang, F. K. Hussain, A. Al-Jumaily, C.-T. Lin, Effects of repetitive ssveps on eeg complexity using multiscale inherent fuzzy entropy, Neurocomputing 389 (2020) 198–206.
4. Z. Cao, C.-T. Lin, Inherent fuzzy entropy for the improvement of eeg complexity evaluation, IEEE Transactions on Fuzzy Systems 26 (2) (2017) 1032–1035.

5. Z. Cao, M. Prasad, C.-T. Lin, Estimation of ssvep-based eeg complexity using inherent fuzzy entropy, in: 2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), IEEE, 2017, pp. 1–5.
6. C. Fu, W. Chang, S. Yang, Multiple criteria group decision making based on group satisfaction, Information Sciences 518 (2020) 309–329.
7. D. Meng, H. Wang, S. Yang, Z. Lv, Z. Hu, Z. Wang, Fault analysis of wind power rolling bearing based on EMD feature extraction, CMES-Computer Modeling in Engineering & Sciences 130 (1) (2022) 543–558.
8. D. Meng, Z. Lv, S. Yang, H. Wang, T. Xie, Z. Wang, A time-varying mechanical structure reliability analysis method based on performance degradation, Structures 34 (2021) 3247–3256.
9. Q. Shang, H. Li, Y. Deng, K. H. Cheong, Compound credibility for conflicting evidence combination: an autoencoder-K-Means approach, IEEE Transactions on Systems, Man, and Cybernetics: Systems (2021) 10.1109/TSMC.2021.3130187doi:10.1109/TSMC.2021.3130187.
10. Y. Deng, Uncertainty measure in evidence theory, SCIENCE CHINA Information Sciences 63 (11) (2020) 210201.
11. M. Zhou, S.-S. Zhu, Y.-W. Chen, J. Wu, E. Herrera-Viedma, A generalized belief entropy with nonspecificity and structural conflict, IEEE Transactions on Systems, Man, and Cybernetics: Systems (2021) DOI: 10.1109/TSMC.2021.3129872.
12. W. Jiang, Y. Cao, X. Deng, A novel Z-network model based on Bayesian network and Z-number, IEEE Transactions on Fuzzy Systems 28 (8) (2020) 1585–1599.
13. Y. Li, D. Pelusi, K. H. Cheong, Y. Deng, The arithmetics of two dimensional belief functions, Applied Intelligence 52 (4) (2022) 4192–4210.
14. D. Li, Y. Deng, K. H. Cheong, Multisource basic probability assignment fusion based on information quality, International Journal of Intelligent Systems 36 (4) (2021) 1851–1875.
15. Z. Wang, Z. Li, R. Wang, F. Nie, X. Li, Large graph clustering with simultaneous spectral embedding and discretization, IEEE Transactions on Pattern Analysis and Machine Intelligence 43 (12) (2021) 4426–4440. doi:10.1109/TPAMI.2020.3002587.
16. Z. Wang, C. Wang, X. Li, C. Gao, X. Li, J. Zhu, Evolutionary markov dynamics for network community detection, IEEE Transactions on Knowledge and Data Engineering (2020) 1–1doi:10.1109/TKDE.2020.2997043.
17. Y. Song, Q. Fu, Y.-F. Wang, X. Wang, Divergence-based cross entropy and uncertainty measures of Atanassov's intuitionistic fuzzy sets with their application in decision making, Applied Soft Computing 84 (2019) 105703.
18. A. P. Dempster, Upper and lower probabilities induced by a multivalued mapping, in: Classic works of the Dempster-Shafer theory of belief functions, Springer, 2008, pp. 57–72.
19. G. Shafer, A mathematical theory of evidence, in: A mathematical theory of evidence, Princeton university press, 1976.
20. R. R. Yager, Generalized Dempster–Shafer structures, IEEE Transactions on Fuzzy Systems 27 (3) (2019) 428–435.
21. L. Xiong, X. Su, H. Qian, Conflicting evidence combination from the perspective of networks, Information Sciences 580 (2021) 408–418.
22. Z. Liu, Y. Liu, J. Dezert, F. Cuzzolin, Evidence combination based on credal belief redistribution for pattern classification, IEEE Transactions on Fuzzy Systems 28 (4) (2020) 618–631.
23. X. Xu, D. Zhang, Y. Bai, L. Chang, J. Li, Evidence reasoning rule-based classifier with uncertainty quantification, Information Sciences 516 (2020) 192–204.

24. Z. Liu, X. Zhang, J. Niu, J. Dezert, Combination of classifiers with different frames of discernment based on belief functions, IEEE Transactions on Fuzzy Systems 29 (7) (2021) 1764–1774.
25. X. Song, F. Xiao, Combining time-series evidence: A complex network model based on a visibility graph and belief entropy, Applied Intelligence (2022) DOI: 10.1007/s10489–021–02956–5.
26. D. Han, J. Dezert, Y. Yang, Belief interval-based distance measures in the theory of belief functions, IEEE Transactions on Systems, Man, and Cybernetics: Systems 48 (6) (2018) 833–850.
27. F. Xiao, J. Wen, W. Pedrycz, Generalized divergence-based decision making method with an application to pattern classification, IEEE Transactions on Knowledge and Data Engineering (2022) DOI: 10.1109/TKDE.2022.3177896.
28. C. Zhu, F. Xiao, Z. Cao, A generalized Rényi divergence for multi-source information fusion with its application in EEG data analysis, Information Sciences (2022) DOI: 10.1016/j.ins.2022.05.012.
29. Z. Wang, F. Xiao, W. Ding, Interval-valued intuitionistic fuzzy Jenson-Shannon divergence and its application in multi-attribute decision making, Applied Intelligence (2022) DOI: 10.1007/s10489–022–03347–0.
30. L. Zhang, F. Xiao, A novel belief $\chi 2$ divergence for multisource information fusion and its application in pattern classification, International Journal of Intelligent Systems (2022) DOI: 10.1002/int.22912.
31. Y. Deng, Random permutation set, International Journal of Computers Communications & Control 17 (1) (2022) 4542. doi:https://doi.org/10.15837/ijccc.2022.1.4542.
32. Y. Deng, Information volume of mass function, International Journal of Computers Communications & Control 15 (6) (2020) 3983.
33. L. Chen, Y. Deng, K. H. Cheong, Probability transformation of mass function: A weighted network method based on the ordered visibility graph, Engineering Applications of Artificial Intelligence 105 (2021) 104438.
34. L. Pan, Y. Deng, An association coefficient of a belief function and its application in a target recognition system, *International Journal of Intelligent Systems 35* (1) (**2020**) 85–104.
35. Bidmc congestive heart failure database, mit-bih normal sinus rhythm database, and long term af database, http://www.physionet.org/physiobank/database Accessed December 5, 2016.