Comparative analysis of three distribution entropy methods for chaos recognition

Xiangjian Zeng¹, Li Wan¹* and Hui Liu¹
¹School of Guangzhou University, Guangzhou, China

*Corresponding author e-mail: wanli@gzhu.edu.cn

Abstract. Distribution entropy (DistEn) is an effective index to measure the complexity of time series. In this paper, the moving distribution entropy (M-DistEn) method is proposed by combining the distribution entropy method with the moving window technology. The moving cut data-distribution entropy (MC-DistEn) method is proposed by combining the DistEn with the moving removal window. Based on the M-DistEn method, moving weighted distribution entropy (MW-DistEn) is proposed. These three distribution entropy methods are used to identify the chaotic state of mixed Logistic map sequences with different parameters. The results show that the M-DistEn method only recognizes the first three mixed states of the sequence, and it cannot accurately recognize the last two chaotic states of the sequence, so it has certain limitations. The MC-DistEn method can accurately identify the four different chaotic states of the sequence, but the DistEn value greatly fluctuates due to the influence of the size of the moving removal window, and the position judgment of sequence state change is not accurate enough. The MW-DistEn method not only accurately recognizes different chaotic states, but also is more accurate in judging the position of sequence state changes and more stable in DistEn value than the M-DistEn method and MC-DistEn method, thereby it has a good application prospect.

1. Introduction

Entropy is a measure of system complexity, and the change of a system state can be described by the change of entropy [1]. Since Shannon put forward information entropy, the theory and application of entropy have developed rapidly, and a series of entropy methods, such as approximate entropy [2], sample entropy [3], and permutation entropy [4], have been proposed successively. However, the commonly used methods of approximate entropy, sample entropy, permutation entropy calculation results depend heavily on embedding dimension and tolerance similar preset parameters, such as, the embedding dimension or similar tolerance exceeding a certain value range will cause the deterioration of the calculation results [5-6]. In order to reduce the influence of the entropy method on preset parameters, Li et al. proposed the distribution entropy [7] (DistEn) in 2015. Because of using the probability density distribution of entropy in the state space vector and the vector distance, make full use of the original sequence is the basis of the information, so distribution entropy is not only dependent on the preset parameter which is small, but also has the high stability and low sensitivity to parameters that these methods do not have [8-9], and the electrocardiographic signal and other fields have been well applied.

Based on the distribution entropy, this paper proposes moving distribution entropy (M-DistEn), moving weighted distribution entropy (MW-DistEn) and moving cut data-distribution entropy
(MC-DistEn). These three methods are applied to chaotic state recognition of sequences. It provides a reference for the practical application of distribution entropy.

2. Methods

2.1. DistEn algorithm
Given a time series \{u(i), i = 1, 2, ... , N\}, DistEn method steps are as follows [7]:

1. Phase space reconstruction: embedding dimension \(m\) and construct vector \(X(i) = \{u(i), u(i+1), \ldots, u(i+m-1)\}\) \((1 \leq i \leq N-m+1)\);
2. Distance matrix construction: define the distance matrix \(D = \{d_{i,j}\}\) among vectors \(X(i)\) and \(X(j)\), \(d_{i,j} = \max\{|u(i+k) - u(j+k)|, 0 \leq i, j \leq N-m-1, 0 \leq k \leq m-1\}\) is the Chebyshev distance between \(X(i)\) and \(X(j)\);
3. Empirical probability density function: the empirical probability density function (ePDF) of distance matrix \(D\) is generated by histogram method. If the histogram has \(M\) bins, we use \(P_h(1 \leq h \leq M)\) to denote the probability of each bin;
4. Calculate the DistEn value: define DistEn as
   \[
   \text{DistEn}(m, M, N) = - \sum_{h=1}^{M} P_h \log_2(P_h);
   \]
5. Calculate the normalized DistEn value: \(\text{DistEn}(m, N) = - \frac{1}{\log_2(M)} \sum_{h=1}^{M} P_h \log_2(P_h)\).

After normalization, the range of DistEn value is \([0, 1]\), which unifies the range of DistEn value with different \(M\) values. Since distance matrix \(D\) is symmetric \((d_{i,j} = d_{j,i})\), in the actual calculation of DistEn value, we can obtain the estimation of ePDF by calculating the \(d_{i,j}\) when \(1 \leq i \leq N-m\) and \(i+1 \leq j \leq N-m+1\), which can reduce the calculation amount and speed up the operation.

2.2. M-DistEn algorithm
M-DistEn algorithm is a kind of system state change recognition method combining moving window technology with DistEn algorithm. Given a time series \(\{u(i), i = 1, 2, ... , N\}\), the steps of M-DistEn algorithm are as follows [10]:

1. Set the size of the moving window as \(W\);
2. Take the moving step \(s\), and start at the left end of the time series \(\{u(i)\}\) and move gradually to the last item, to get the intercell \(\text{int}[1+(N-W)/s]\) (int stands for integer);
3. Calculate the DistEn value between each intercell to get the DistEn value sequence;
4. Observe the DistEn value sequence changes to determine the state changes of the sequence.

2.3. MC-DistEn algorithm
Given a time series \(\{u(i), i=1, 2, ..., N\}\), MC-DistEn algorithm steps are as follows [11]:

1. Set the window length of moving removed data as \(W\);
2. Removing data with length \(W\) continuously from time series \(\{u(i)\}\), connecting the remaining \(N-W\) data, then a new subsequence of \(N-W\) is obtained;
3. Calculate the DistEn value of the new subsequence by using the DistEn algorithm;
4. Taking the moving step \(s\) and moving the window gradually by keeping the window scale of data removed unchanged. Repeat steps (2) and (3) to get a DistEn value sequence with length \(\text{int}(N/M)\) (int stands for integer);
5. Observe the DistEn value sequence changes to determine the state changes of the sequence.

2.4. MW-DistEn algorithm
In consideration of the fact that DistEn algorithm of histogram method is used to generate probability density function is a reflection of the number of data, in order to highlight the characteristics of the data itself and make full use of the degree of discrete information in the data, in the process of calculating DistEn value we used probability distribution in each group in the space vector distance
variance as the weight. The weighted distribution entropy (W-DistEn), set \( \text{Var}_h \) \( (1 \leq h \leq M) \) histogram of variance, the first group \( h \) data is W-DistEn formula is defined as:

\[
W-\text{DistEn}(m, N) = -\frac{1}{\log_2(M)} \sum_{h=1}^{M} \text{Var}_h P_h \log_2(P_h).
\]

Given a time-series \( \{u(i), i = 1, 2, \cdots, N\} \), steps (1) and (2) of MW DistEn algorithm are the same as those of M-DistEn algorithm. Instead of calculating the DistEn value in step (3) of the M-DistEn algorithm, we calculate the W-DistEn value and get the W-DistEn value sequence. Finally, we observe the change of the W-DistEn value sequence to determine the state change of the sequence.

3. Chaotic state recognition

3.1. Construct chaotic sequence IS1

Logistic mapping is the most common mathematical model to describe the evolution of a biological population system. The equation is as follows: \( x_{n+1} = \lambda x_n (1 - x_n) \), \( x_n \in [0,1] \). When \( 3 \leq \lambda \leq 4 \), the system leads to chaos by times period. When the parameter reaches the limit value \( \lambda_{\infty} = 3.569945673 \), the final state of Logistic mapping iterative time series is period 2\(^\infty\), that is, the system enters a chaos state at this time [12], as shown in Figure 1(a).

![Figure 1](image)

**Figure 1.** The bifurcation diagram of Logistic map

In order to compare the application of M-DistEn, MW-DistEn and MC-DistEn in chaotic state recognition, the ideal sequence IS1 generated by Logistic mapping is constructed. Taking the initial value \( x_0 = 0.4 \), the first 2000 data parameter \( \lambda_1 = 3.57 \) for the ideal sequence IS1, the 2001 to 4000 data parameter \( \lambda_2 = 3.62 \), the 4001 to 6000 data parameter \( \lambda_3 = 3.67 \), the 6001 to 8000 data parameter \( \lambda_4 = 3.72 \), and the total length is 8000. Sequence IS1 is composed of four data segments with different parameters, each of which has a length of 2000, and different parameters correspond to different chaotic states respectively. The system evolves from the lower growth \( \lambda_1 \) to the higher \( \lambda_2 \), then to the higher \( \lambda_3 \), and finally to the highest \( \lambda_4 \), parameter \( \lambda_1 < \lambda_2 < \lambda_3 < \lambda_4 \), as shown in Figure 1(b).

3.2. M-DistEn detection results of ideal sequence IS1

The M-DistEn method was used to calculate the sequence IS1. The embedding dimension takes the usual \( m = 2 \), the size of the moving window \( W = 400 \), and the moving step \( s \) is taken as 10, 20, 30 and 50 respectively, and the results are shown in Figure 2.
Figure 2. M-DistEn detection results of ideal time-series IS1

The M-DistEn detection results of IS1 in moving step size for $s = 10, 20, 30, 50$ are shown in Figure 2. For the four different moving steps, the DistEn value variation trend obtained by the M-DistEn method is almost the same, at $n=2001$ and $n=4001$, the DistEn value increases significantly. It shows that the complexity of sequence increases; and the M-DistEn method recognizes three different chaotic states in the first 6000 data in the sequence. At $n=6001$, although the DistEn value fluctuates slightly, the DistEn value of the data 4001-6000 and 6001-8000 are close to 1, which is close to the upper limit of the DistEn value. It is impossible to determine whether the sequence is in two different chaotic states. It indicates that the M-DistEn method cannot recognize the two chaotic states between data 4001-6000 and data 6001-8000. Therefore, the M-DistEn method can only accurately identify the first three chaotic states of the sequence IS1, and cannot accurately identify the latter two chaotic states with DistEn values close to 1.

3.3. MC-DistEn detection results of ideal sequence IS1

The MC-DistEn method is used to calculate the sequence IS1. Take the embedding dimension $m=2$, the moving step $s=50$, and the moving removal window take 80, 160, 240 and 400 respectively, corresponding to 1%, 2%, 3% and 5% of the original sequence length, and the results are shown in Figure 3.

As can be seen from Figure 3, the MC-DistEn detection results obtained by the four moving removal windows are almost the same. The DistEn value changes similarly under different moving removal windows. At $n=2001$, $n=4001$ and $n=6001$, the DistEn value drops significantly, indicating that the complexity of the removed data is getting higher and higher, leading to a downward trend in the complexity of the remaining data. It indicates the complexity of data corresponding to parameter $\lambda_1$ in IS1 sequence < the complexity of data for the parameter $\lambda_2$ < the complexity of the data for the parameter $\lambda_3$ < the complexity of the data corresponding to parameter $\lambda_4$. It indicates that the MC-DistEn method can accurately recognize the four different chaotic states in the sequence IS1, which is better than the M-DistEn method. However, MC-DistEn methods among different levels before and after DistEn values exist a similar length and the moving remove window size change interval, as the moving removal window increases, so do the range of change. It is difficult to accurately determine the starting position of the range of change, and the DistEn value corresponding to different moving removal windows fluctuates by different extents. The smaller the move removal window, the more DistEn fluctuates.
Finally, the MW-DistEn method is used to calculate the sequence IS1. The embedding dimension $m=2$, the size of the moving window $W=400$, and the moving step $s$ is 10, 20, 30, and 50, respectively. The results are shown in Figure 4.

As can be seen from Figure 4, the variation trend of W-DistEn values obtained by the four moving steps is almost the same. At $n=2001$, $n=4001$ and $n=6001$, the W-DistEn value increases, indicating that the position of sequence state changes is at $n=2001$, $n=4001$ and $n=6001$, respectively. With the increase of parameters, the W-DistEn value becomes larger,
that is, the complexity of the sequence becomes higher, which indicates that the MW-DistEn method can accurately identify the four chaotic states in the sequence IS1 and accurately judge the position of state changes. In addition, the W-DistEn value of different states fluctuates less and has good stability. Therefore, the MW-DistEn method is better than the M-DistEn method and the MC-DistEn method in the recognition of sequence chaotic states.

4. Conclusion
In this paper, three distribution entropy methods, M-DistEn, MC-DistEn and MW-DistEn, are proposed based on DistEn. Based on Logistic mapping, the mixed sequence IS1 with different parameters was constructed. Three distribution entropy methods are utilized to test the effect of a different chaotic state. The results show that: M-DistEn cannot fully recognize the four chaotic states in the sequence, and when the DistEn value is close to the upper limit of the DistEn value, its ability to recognize the chaotic states of the sequence is greatly weakened, which has certain limitations; MC-DistEn makes up the deficiency of M-DistEn. It can distinguish the four chaotic states of the sequence, but its detection results are not stable and accurate. The MW-DistEn method can not only accurately recognize different chaotic states, but also accurately judge the position of state changes. Besides, the W-DistEn value in different states fluctuates very little and has good stability. The recognition effect of the MW-DistEn method is better than that of M-DistEn method and MC-DistEn method, thus it has a good application prospect.

5. Acknowledgments
This research was financially supported by the National Science Foundation of China (41872246); Innovation Research for the Postgraduate of Guangzhou University (2019GDJC-M29).

References
[1] Kaffash F , Foglyano R , Wilson C G , et al. The effect of time delay on Approximate & Sample Entropy calculations[J]. Physica D Nonlinear Phenomena, 2008, 237(23):3069-3074.
[2] Steven M. Pincus. Approximate Entropy as a Measure of System Complexity[J]. Proceedings of the National Academy of Sciences of the United States of America, 1991, 88(6) 2297-2301.
[3] Richman J S, Moorman J R. Physiological time-series analysis using approximate entropy and sample entropy.[J]. American journal of physiology. Heart and circulatory physiology, 2000, 278(6) 2039-2049.
[4] Bandt C , Pompe B . Permutation Entropy: A Natural Complexity Measure for Time Series[J]. Physical Review Letters, 2002, 88(17):174102-0.
[5] P. Castiglioni, M. Di Rienzo. How the threshold "r" influences approximate entropy analysis of heart-rate variability[C]/ Computers in Cardiology. IEEE, 2008 561 - 564..
[6] Yentes J M , Hunt N , Schmid K K , et al. The Appropriate Use of Approximate Entropy and Sample Entropy with Short Data Sets[J]. Annals of Biomedical Engineering, 2013, 1(2):349-365.
[7] Li P , Liu C , Li K , et al. Assessing the complexity of short-term heartbeat interval series by distribution entropy[J]. Medical & Biological Engineering & Computing, 2015, 53(1):77-87.
[8] Flanagan R , Lacasa L . Irreversibility of financial time series: A graph-theoretical approach[J]. Physics Letters A, 2016, 380(20):1689-1697.
[9] Sprott J C , Rowlands G . Improved correlation dimension calculation [J]. International Journal of Bifurcation & Chaos, 2001, 11(07):1865-1880.
[10] Sun D , Huang Q , Wang Y , et al. Application of moving approximate entropy to mutation analysis of runoff time series[J]. Journal of Hydroelectric Engineering, 2014, 33(4):1-6.
[11] H. M. Jin, W. P. He, W. Zhang, A. X. Feng, and W. Hou, “Effect of noises on moving cut data-approximate entropy,” Acta Phys. Sin., Vol. 61, No. 12, 2012, pp. 069201-160.
[12] W. F. Shi, Research on Logistic Map and Its Chaotic Characteristics[J]. Electro-optic Technology Application, 2004,19(2):53-56