2011 International Conference on Computer Communication and Management
Proc .of CSIT vol.5 (2011) © (2011) IACSIT Press, Singapore

Distribution Probability Matrix Algorithms of Mutual Information

Ai-Hua JIANG [1,*], Zhen-Hua ZHANG [1], Xiu-Chang HUANG [1]

[1] State Key Laboratory of Mechanical System and Vibration, Shanghai Jiao Tong University, No.800DongchuanRoad, Shanghai200240, China.

Abstract: Veracity and rate of distribution probability matrix algorithms are critical for statistical mutual information algorithms, which are effective for time delay in phase space reconstruction process. Firstly, two categories of mutual information algorithms are summarized following the partition criterions of plane, which is constructed by a pair of series with the same size. And then two algorithms of distribution probability matrix that shows the distribution of points corresponding to the data pairs of Lorenz series on the plane are raised based on sorting the two series and replacing each numerical value with its order number in its own series so as to judge the element in which data sets are located or count point amount in certain element. The optimal time delay of the two distribution probability matrix algorithms as well as the computing time is also compared when series sizes are different. The results show that the time consumption of distribution probability matrix algorithm by judging the element in which data sets are located increases linearly and is much less than that of algorithm by counting point number in certain element, but both algorithms can reach the same distribution probability matrix and optimal time delay.

Keywords: Mutual information algorithm, Distribution probability matrix, Time delay

1. Introduction

1.1. Theory of mutual information

Mutual information is widely used to ascertain one of the most important parameters, time delay, in reconstructing phase space from nonlinear time series. And Fraser upraised the first mutual information algorithm[1].

Provided that $\{s(t_i)\}$ (i=1,2…N) is a nonlinear series which comes from experiment and whose data acquiring interval is Δt , the series can be reconstructed as bellow.

$(s(t_1), s(t_1+\tau),\ldots, s(t_1+(d-1)\tau))$,

$(s(t_2), s(t_2+\tau),\ldots, s(t_2+(d-1)\tau))$,

……

$(s(t_N), s(t_N+\tau),\ldots, s(t_N+(d-1)\tau))$.

Where τ represents time delay. And the mutual information between $S=\{ s(t_1), s(t_2),\ldots, s(t_N)\}$ and $Q=\{ s(t_1+\tau), s(t_2+\tau),\ldots, s(t_N+\tau)\}$, I(S,Q), is the average bits that S can be predicted by the measurement from Q. I(S,Q) can be expressed as:

$$I(S,Q)=H(Q)+H(S)-H(S,Q) \quad (1)$$

Where H(Q) and H(S) are the entropy of Q and S respectively, and H(S,Q) is the mutual entropy between S and Q. Normally, the moment of the first minimal mutual information is taken as the optimal time delay for phase space reconstruction.

To figure out the mutual information between two time series, numerical values of S and Q are generally corresponded to the points of two perpendicular axes respectively. A plane, SQ, is also introduced in this process and data pairs of S and Q are corresponding to the points in SQ plane. Then the interval between the maximal value of Q and the minimal value of Q is partitioned to a certain number of elements, $\{q_i\}$, (i=1,

*Corresponding author. PH.D. Candidate of State Key Laboratory of Mechanical System and Vibration, Shanghai Jiao Tong University, Shanghai, China. Tel.:84+21+34206095;
E-mail addresses:
woshiaihua@sjtu.edu.cn.

2…), according to different standards. The sketch map of partition on SQ plane is shown as figure 1. And the probability distribution of the points corresponding to data pairs in q_i, $P(q_i)$, can be gained by dividing the size of data series , N. Furthermore, the partitioning of axis S is similar to that of axis Q. In addition, the superposition between s_i and q_j can be expressed as (s_i,q_j) and the probability distribution in (s_i,q_j), $P(s_i,q_j)$, is the ratio between the point amount in (s_i,q_j) and N. Thus, the mutual information between S and Q can be calculated by equation (1), and $H(S)=-\sum P(s_i)\log_2 P(s_i)$, $H(Q)=-\sum P(q_i)\log_2 P(q_i)$.

$$H(Q,S)=H(S,Q)=-\sum_{i,j} P\left(s_i,q_j\right)\log_2 P\left(s_i,q_j\right)$$

Where $P(s_i)$ is the probability distribution of section s_i. And equation (1) changes to equation (2) as well.

$$I(Q,S)=\sum_i\sum_j P_{sq}(s_i,q_j)\log_2\left[\frac{P_{sq}(s_i,q_j)}{P_s(s_i)P_q(q_j)}\right] \quad (2)$$

1.2. Criterions to partition SQ plane

a) Equal probability distribution b) Equal distance

Fig. 1 Sketch map of Partitioning SQ plane

There are mainly two criterions, namely equal probability distribution in every s_i or q_j and equal distance between the upper limit and lower limit of each s_i or q_j, used for the partition of SQ plane. Fig. 1(a) shows the partition of SQ plane based on the first standard. It can be seen that the amount of the points located in each element, s_i or q_j, is the same, but the widths of elements may be different. Fig. 1(b) displays the partition of SQ plane based on the second criteria. In Fig. 1(b), no matter how many points in s_i or q_j, the width of each element is the same.

Cellucci putted forward a statistic mutual information algorithm based on equal probability distribution [2], and Jiang improved this algorithm by simplifying the algorithm of judgment of substructure in each element [3]. The latest algorithm was constructed based on the null hypothesis that series S and series Q are statistically independent. If the hypothesis is true, points of data pairs coming from S and Q distribute uniformly in SQ plane. Then mutual information computed by equation (2) is zero and the amount of points in each grid is known. Supposing that $O(s_i)$ is the quantity of points in ith element on axis S and $O(q_j)$ is the quantity of points in jth element on axis Q, the amount of points in (i,j) should be a known number, $E(s_i,q_j)$, which is shown in equation (3).

$$E(s_i,q_j)=NP(s_i)P(q_j)=N\left(\frac{O(s_i)}{N}\right)\left(\frac{O(q_j)}{N}\right)=\frac{O(s_i)O(q_j)}{N} \quad (3)$$

When the standard of equal probability distribution is used to partition the SQ plane and both element numbers of S and Q are N_E, then $P(s_i)$ and $P(q_j)$, the probability of s_i and q_j respectively, are $P(s_i)= P(q_j) =1/N_E$. So equation (3) becomes equation (4).

$$E(s_{i,}q_j)=NP(s_i)P(q_j)=\frac{N}{N_E^2} \quad (4)$$

Here if $E(s_i,q_j)$ is a known number, the element number N_E should be also a known number from equation (4). In Cellucci's algorithm, $E(s_i,q_j)\geq 5$ was taken because there is little probability that substructures exist in grid under this condition. And equation (4) changes to equation (5).

$$N_E\leq\left(\frac{N}{5}\right)^{1/2} \quad (5)$$

Normally, N_E is taken as the maximal integer which satisfies equation (5) in partitioning SQ plane. Then each element has the same number of points, and the equation (2) becomes equation (6).

$$I(Q,S)=\sum_{i=1}^{N_E}\sum_{j=1}^{N_E}P(s_i,q_j)\log_2\left(N_E^2P(s_i,q_j)\right) \quad (6)$$

And the maximal integer number, N_1, which is equal to or less than N/N_E can be taken as number of data points in each s_i or q_j to partition axis S and axis Q from their minimal value to maximal value respectively. If N/N_1 is an integer, the partitioned elements have the same data points, whereas N/N_1 is not an integer, the last element of axis S and the last element of axis Q have data points less than N_1. Then, mutual information

can be directly computed from original probability matrix in the first case, and there are three methods to dispose probability matrix in the second case.

The first method is to calculate mutual information from matrix which removes the last row and the last column of the original probability matrix when N/N_1 is not an integer. Because Cellucci algorithm is promoted based on the statistic of a great deal of data pairs, ignoring limit data pairs will indistinctively affect the result and the max amount of data points to be neglected is just N_1-1.

Another method is to calculate mutual information by the original probability matrix, no matter whether the N/N_E is an integer or not. Taking five as the anticipant data points in (s_i,q_i) is by the reason of no substructures existing in grids in that case, and the data points in the grids of the last row and the last column are less than five when N/N_1 is not an integer, thus acquiring mutual information right from original probability matrix is also rational in this aspect. And the original probability matrix is expressed as C_2.

The last method is to modify the last row and the last column of C_2 according to the scale between data points in the last element and data points in other element and idiographic mean is as following. Provided N_2 is the residue when N divides N_E, then a new matrix, C_3, appears by multiplying the occupancies of the last row and the last column in C_2 with N_1/N_2 respectively. What need to be pointed out is that the superstition of the last row and the last column in C_2 should multiply N_1/N_2 once.

2. Distribution probability matrix algorithms

2.1. Judging the element which data sets locate in

It can be known from the algorithm described above that the key step in mutual information calculation is how to gain the probability distribution matrix. The time consumption and space consumption in that step occupy the majority of the whole mutual information algorithm.

In the case of partition SQ plane with equal probability on both two axes, the matrix only relates to the order of each numerical value in series, but not the numerical value itself, So any changes on the numerical value of series are acceptable as long as the order of each number in series keeps invariable. This trait gives birth to rapid and easily realized algorithms.

Estimation of the original probability distribution matrix can base on the following method. Above all, the two series, S and Q, are sorted respectively by a rapid sorting program, and then the numerical values of S and Q are replaced by their order numbers respectively. As a result, two new data vectors, A and B, are obtained from S and Q respectively, and no matter how large the numerical ranges of S and Q are, numerical values in A and B are integer from 1 to N. Therefore, eliciting probability matrix from A and B is much easier than from S and Q. Due to element data amount, N_1, is a known number, the grid that data set, (A(i),B(i))（i=1,2…N）, falls into is also clear. For each step to judge the position of (A(i),B(i)), the probability matrix occupancy, C(m,n), is added one. Where m is the maximal integer which is less than $1+A(i)/N_1$ and n is the maximal integer which is less than $1+B(i)/N_1$. Following this course, the original probability, C_2, matrix can be obtained.

2.2. Counting point number in certain element

In this case, the numerical values of S and Q are also replaced respectively by their order numbers before the matrix calculation in order to compare the time of two algorithms.

So the values in vectors, A and B, are also from 1 to N. And then A(i) , where i is from 1 to N, is seriatim picked out as target vale and there is also a value in vector B equal to A(i). By finding the value B(j) equal to A(i), the index j is noted. At last, the value of original probability matrix, C(m, n), is added one. Where m is the maximal integer which is less than 1 to $1+i/N_1$ and n is the maximal integer which is less than $1+j/N_1$. After all the vales in A finding its equal number in B, t he original probability matrix C_2 is also computed.

After C_2 is acquired, equation (2) changes to equation (7) based on the first method raised in section 1.2.

$$I(Q,S)=\sum_{m=1}^{k}\sum_{n=1}^{k}\frac{C(m,n)}{N^*}\log_2\left[\frac{C(m,n)\cdot N^*}{N_1^2}\right] \qquad (7)$$

Where N^* equals to $N_1\cdot k$, and k is the maximal integer equal to or less than N/N_1. If the second method described in 1.2 is used, equation (2) changes as to equation (8).

$$I(Q,S)=\sum_{m=1}^{k+1}\sum_{n=1}^{k+1}\frac{C(m,n)}{N}\log_2\left[\frac{C(m,n)\cdot N}{N(m,n)}\right] \qquad (8)$$

Where $N(m,n)=N_1\cdot N_2, m=k+1 \ \ or \ \ n=k+1; N(m,n)=N_2^2, \ \ m=n=k+1; N(m,n)=N_1^2, \ \ others$；

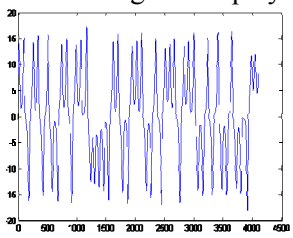
If the third method described in1.2 is used, equation (2) changes to equation (9).

$$I(Q,S)=\sum_{m=1}^{k+1}\sum_{n=1}^{k+1}\frac{C^*(m,n)}{(k+1)\cdot N_1}\log_2\left[\frac{C^*(m,n)\cdot(k+1)}{N_1}\right] \quad (9)$$

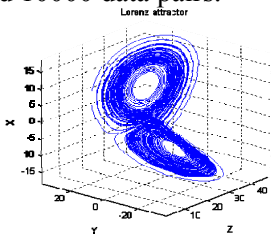
Where $C^*(m,n)=C(m,n),0<m,n\le k$, $C^*(m,n)=\frac{N_1}{N}C(m,n),others$.

3. Validation of the two distribution probability matrix algorithms

To compare and validate the two algorithms of probability matrix, series of Lorenz equation $dx/dt=\sigma(x-y)$, $dy/dt=-xz+rx-y$ and $dz/dt=xy-bz$, where σ =10, b=8/3 and r=28, are taken as an example. One hundred thousand data pairs, which originate from (8.331,13.291,18.063), are obtained by four-order Runge-Kutta algorithm with 0.01-second time step. And all the process to generate Lorenz data pairs can be carried out by Tisean3.0.0 [4]. Figure 2(a) is the first 4096 data points of x in Lorenz equation and 2(b) is the three-dimensional diagram displaying the forward 10000 data pairs.



a）The former 4096data points of x



b）10000 data pairs of x, y and z

Fig.2 Time Series of Lorenz

3.1. Optimal time delay by two algorithms

According to the two distribution probability algorithms formatted in section 2, time series with different lengths in Lorenz system are used to calculate the optimal time delay. The forward 1024, 2048, 4096, 8192, 16384, 24576, 32768 and 65536 data points of variable x are utilized respectively as the first series, S, and another series which lag S from 1 to 200 in the ten thousand data point of x are used as the second series, Q. So for each length of time series, 200 mutual information values can be acquired.

And with the two algorithms showed in 2.1, multi groups of mutual information series are gained, and the location of the first minimal value in mutual information series indicates the optimal time delay.

Both the two probability algorithms are feasible, and the original probability matrixes are the same by the two different algorithms, so as the mutual information with the same series. Fig 3 shows the mutual information of the two algorithms with different series lengths.

Table 1 shows the accurate parameters appearing in the process of mutual information calculation by series with different lengths listed above. Where the other symbols are the same as those referred above.

3.2. Time consumption of two mutual information algorithms

It can be seen from table 1 that following the increase of the length of data pairs, the optimal time delays, τ , i.e. the lag of the first minimal mutual information, are stabilizing no matter which probability distribution matrix is employed, and these optimal time delays fix at 17, which is much close to the theory optimal time delay of Lorenz system[5,6], i.e. 18, when the length of time series is equal to or bigger than 16384.Time consumption of two mutual information algorithms.

Table 1 also shows the time consumption of the two algorithms. It can be seen that both algorithms spend more time to compute the mutual information of two series following increase of the data length, N.

The algorithm by counting point number in certain element to compute probability matrix consumes much more time than that of the other algorithms. When data length is only 2048, the algorithm with counting point in certain element costs more than 35 seconds, which is much longer than 15 seconds, time consumption of the other algorithm with 65536 data sets. Furthermore, when data length is not shorter than 8192, the algorithm by counting point number in certain element spends more than 1.5 minutes. All the algorithms are realized by Matlab6.5 installed on a PC with 3GHz CPU frequency and 2GB memory.

Figure 4 shows the relationship between the data length and time consumption of the two algorithms. It can be known that following the increase of N time of algorithm by judging the element which data sets locate in spent on the calculations enhances with nearly linear trend, while time of the other algorithm upraises with exponential trend.

Table 1 Parameters appearing in the process of mutual information calculation by series with different lengths

N	E_{SQ}	N_E	N_I	N/N_E is integer	Judging Data Sets' Element						Counting Point Number		
					C_1		C_2		C_3		$C1$	$C2$	$C3$
					Time Delay	t (s)	Time Delay	t (s)	Time Delay	t (s)	t (s)	t (s)	t (s)
1024	5	14	73	No	22	0.19	17	0.172	16	0.19	2.375	2.36	2.381
2048	5	20	102	No	20	0.38	17	0.375	17	0.39	9.109	9.078	9.156
4096	5	28	146	No	19	0.78	19	0.765	19	0.8	35.48	35.33	35.88
8192	5	40	204	No	18	1.61	18	1.594	18	1.64			
16384	5	57	287	No	17	3.38	17	3.344	17	3.44			
24576	5	70	351	No	17	5.22	17	5.187	17	5.3			
32768	5	80	409	No	17	7.06	17	7.062	17	7.19			
65536	5	114	574	No	17	15.6	17	15.58	17	15.9			

It also can be known from Figure 3 and Table2 that, time of each algorithm calculation with C_1, C_2 and C_3 is basically adjacent, and the time with C_2 is delicately less than the others.

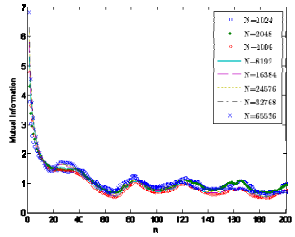


Fig. 3 Mutual information of both algorithms with different series lengths

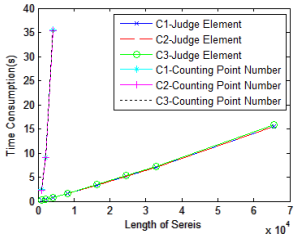


Fig.4 Relationship between length of time series and time consumption

4. Conclusion

This paper compared two probability distribution algorithms of statistic mutual information. The following results are obtained in the compare of these algorithms:

1) Both the probability distribution algorithm by judging the element which data sets locate in and the one by counting point number in certain element are available and feasible to calculate mutual information based on the statistic mutual information.

2) The time consumption of the probability distribution algorithm by judging the element which data sets locate in increases linearly following the enhancement of data sets length, as time consumption of the algorithm by counting point number in certain element increases with exponential trend.

5. Acknowledgements

The discussions about usage of Tisean 3.0.1 with Lei Min, assistant professor of Shanghai Jiao Tong University, are acknowledged with gratitude.

6. References

[1] A. M. Fraser, H. L. Swinney, Independent coordinates for strange attractors from mutual information. Physical Review A, 33 (1986)1134 - 1140.

[2] C. J. Cellucci, A. M. Albano, P. E. Rapp, Statistical validation of mutual Information calculations: comparisons of alternative numerical algorithms, Physical Review E, 71(2005) 1-14.

[3] Ai-Hua Jiang, Xiu-Chang Huang, Zhen-Hua Zhang, etc. Mutual information algorithms. Mechanical Systems and Signal Processing. 24 (2010) 2947–2960.

[4] R. Hegger, H. Kantz, and T. Schreiber, Practical implementation of nonlinear time series methods: The TISEAN package, CHAOS 9, (1999) 413-435.

[5] Edward Ott, Chaos in Dynamic System, Cambridge University Press, Cambridge, 2002

[6] H. Kantz, T. Schreiber, Nonlinear Time Series Analysis, Cambridge University Press, Cambridge, 2004.