International Journal of Computer Applications (0975 8887)
Volume 93 - No. 15, May 2014

Different Methods for Partitioning the Phase Space of a Dynamic System

Abir Hadriche, Nawel Jmail
Laboratory of Electronics and Information Technologies
Sfax University, Sfax, Tunisia.

Ridha Elleuch
Laboratory of Géoressouces, Matériaux, Environnement.
Physics Department
Faculty of Sciences of Sfax
Sfax University, Sfax, Tunisia.

Laurent Pezard
Institut de Neurosciences des Systèmes (UMR S 1106)
Aix-Marseille University, INSERM, Marseille, France.

ABSTRACT

In symbolic dynamics, the definition of a symbolic sequence from a continuous times series depends on the use of an appropriate partition of the phase space. In fact, the best way is to estimate a generating partition.
However, it is not possible to find generating partitions for most experimental observations because such partitions do not exist when noise is present.
In this paper, different partition methods applied to stochastic and chaotic system will be compared in order to choose one which conserves system entropy rate. This partition is called a Markov partition.

General Terms:

Markov partition, symbolic dynamic

Keywords:

Markov partition, generating partition, symbolic sequence, entropy rate

1. INTRODUCTION

Symbolic dynamics approach allows to study a discrete dynamical system equivalent to the continuous dynamics of a physical system. This approach was pioneered by Hadamard in 1898 [12] who suggested the idea to split the state space into a finite number of parts, each part having a name (usually a number or letter of alphabet). All the points of the phase space are given the name of the part to which they belong.
Given a partition of the state space, associating to each point x an infinite word describing its trajectory is a transformation from a complex application to a simple one. But the state space becomes more complex, it is now a set of infinite words which allows to understand the structure of the dynamics.
The main issue of symbolic dynamics is to seek for a partition of the state space that describes the trajectories of points and verify the constraint to represent the dynamics in a simple way. This symbolic method allows to overcome the limitations of analytical approaches, while retaining some key properties in terms of the dynamics.
This type of partition is called a Markov partition because of its connection to discrete time Markov processes.
In the dynamical systems literature, one typically search for a generating partition which has strong theoretical settings. Nevertheless, a generating partition is hardly determined on the basis of experimental data [5].
Note further that every Markov partition is generating, but the converse is not necessarily true [5].
In this paper, different methods for partitioning the phase space of a dynamic system will be presented such as methods of vector quantification (self organizing maps [16] and K-means algorithm [20]) and methods of numerical approximation of dynamical systems (varcluster [1] and subdivision [7]).
These methods will be tested in chaotic and stochastic dynamic systems to extract the best partition associated to Markov process.
This paper is organized as follows: the first section describes various phase space partition methods, these methods will be used in the following section to quantify discrete stochastic and chaotic systems in order to estimate their performance and select the best one.

2. METHODS

2.1 Different partition methods

In this paper, the phase space of each dynamical system is denoted $\mathcal{E}$.
In this space, a simulation of T_s seconds leads to a sequence of row vectors $\vec{e}(t) = (e_1(t), \ldots, e_N(t))$ with $t = 0, \ldots, T-1$ with $T = T_s/\Delta t$ the number of samples simulated by Δt integration step. The partition of $\mathcal{E}$ corresponds to the definition of non-overlapping regions m_i so that $\bigcup_i m_i = \mathcal{E}$ and $\bigcap_i m_i = \emptyset$. This procedure leads to coarse grained the $\mathcal{E}$ space and to obtain a new $\mathcal{M}$ space with $||\mathcal{M}|| = m$ regions.

2.1.1 First method: K-means. This method of clustering was introduced by MacQueen [20] and its algorithm is developed by Hartigan and Wong 1979 [14]. This method is the simplest one that can solve the problems of partitioning. Its seeks the easiest way to classify the data in some m numbers region in state space [6], [21]. The idea is to represent each group by its mean centroid called c. The steps of this algorithm are:

—Choose randomly m groups.

—Regroup centroids that are close enough.

—Recalculate the positions of the m new centroids.

—Repeat last two steps until the minimization of measurement error given by the sum of square errors calculated between each point and each group centroid.

$$E = \sum_{j=1}^{m} \sum_{i} \|x_i^j - c_j\|^2 \tag{1}$$

With $\|x_i^j - c_j\|^2$ is the distance measured between x_i^j points included in j^{th} group with c_j centroid.

2.1.2 Second method: Self organizing maps. Unlike the k-means method, the self-organizing maps (SOM) [16] allows a rapid unsupervised learning of individuals (the states of systems in phase space).
It relies on a neural network distributed uniformly in a space of 2 or 3 dimensions. Each neuron is defined by a vector in the space of individuals, called weight vector. Individuals are presented successively to the network.
For each individual x_k with $k \in \{1, ..., T\}$, the nearest neuron (called *Best Matching Unit, BMU*) and its vicinity in the network are modified so that together they are close to the individual.
This algorithm takes place mainly in three phases:

—Initialization the weights of output neurons (small random values).

—Presentation of an example of the base and determining the output neuron closest to the example (*BMU*).
Finally, we determine the Euclidean distance between the example and all the output neurons characterized by their weight. The neuron i is selected:

$$\|w_i - x_k\| \leq \|w_j - x_k\| \quad \forall j \neq i \tag{2}$$

Euclidean distance is used to calculate the activation of each output neuron as follows:

$$\alpha_j = \frac{A}{B + C\|w_j - x_k\|} \tag{3}$$

With x_k is the input of the map, w_i is the weight of the neuron and A, B, and C are any constants.

—Weights are adapted using:

$$w_j(t+1) = w_j(t) + \alpha(t)v(j,i,t)(x - w_j(t)) \tag{4}$$

$\alpha(t)$ is learning rate and $v(j,i,t)$ is neighbors function. Only the weights of neurons in the vicinity of the selected one are changed. Learning rate $\alpha(t)$ is a function that decreases over time:

$$\alpha(t) = \frac{\alpha_0}{1 + K_\alpha t} \tag{5}$$

The neighbors function $v(j,i,t)$ is a Gaussian function that evolves such as :

$$v(j,i,t) = \exp\left(-\frac{d^2(j,i)}{2\sigma^2(t)}\right) \tag{6}$$

This is repeated until ϵ reaches a threshold. ϵ is defined as:

$$\epsilon = \frac{\sum_{i=1}^{n} \|w_{k_i} - xi\|^2}{n} \tag{7}$$

for n iterations.

2.1.3 Third method: Subdivision. This partitioning method is introduced by Hohmann and Dellnitz [7] inspired by Ulam-Galerkin discretization process [22]. It was used for the quantification of random dynamical systems (stochastic oscillations of Van Der Pol) [15], and also of chaotic systems (logistic and Hénon applications) [8].
We start with a dynamic system given by the application $f : R^{\rightarrow} R^N$, the algorithm considers the partition in the first place a rectangle Q given by:

$$Q = R(c,r) := \{x = (x_i) \in R^N : |x_i - c_i| \leq r_i \text{for } i = 1, \ldots, N\} \tag{8}$$

With $c = (c_i)$, $r = r_i \in R^d$ and a space phase of the system given by:

$$A_Q := \bigcap_{n \geq 0} f^n(Q) \tag{9}$$

and ignoring any dynamic outside the rectangle Q.
The selection of regions is valid in the rejection of all empty rectangles Q that doesn't contain an image of f (x_i in phase space).

2.1.4 Fourth method: Varcluster. This approach of partition of the state space is proposed by Allefeld et al. [1].
They defined the state space by N variables [1] $(x_1, x_2, \ldots, x_N) = x$ to be discretized, resulting in a set of compound microstates which forms the basis for further analysis.
This algorithm uses a recursive bi-partitioning approach: for a given set of T data points $M = \{x(t)\}$, $t = 1 \ldots T$, the direction of maximal variance is determined, i.e a unit vector e, $|e| = 1$, such that $var_m\,(x(t).e)$ obtains its maximum value. Using the median Med of the data points' positions along this direction as a threshold value, the set is divided into two subsets:

$$M_1 = \{x(t) | x(t).e \leq Med\} \tag{10}$$

$$M_2 = \{x(t) | x(t).e > Med\} \tag{11}$$

The procedure is repeated for each of the resulting subsets. The number of repetition is the number k of iteration to obtain a sequence of 2^k symbols.

2.1.5 Comparaison between different partition methods. To extract the best partition methods to describe the dynamic system in phase space, we used this strategy:

—Divide the phase space into $m = 2^k$ regions using one of the partition methods explained above.

—**Calculate transition matrix of the Markov chain**:
The dynamics of Markov chain of order 1 is fully determined by the transition matrix τ and stationary distribution π. Then for m space regions, a matrix of [m x m] dimensions was established (similar to a transfer matrix in statistical physics), whose transi-

[1] N is the state's dimension

tion elements are in general either real or complex weights.

$$\tau_{i,j} = \begin{cases} \tau_i \in R \text{ or } C & \text{if the transition } m_j \to m_i \text{ is permitted,} \\ 0 & \text{else.} \end{cases} \tag{12}$$

Here, let $\mathcal{M}^{(k)}$ be an homogeneous Markov chain for partition k, its transition matrix $\tau^{(k)} = \tau_{ij}^{(k)}$ with $i, j \in \{1 \dots 2^k\}$; so the transition probability between states is $\tau^{(k)} = (\tau_{i,j}^{(k)}) = (\Pr(m^{(k)}(t) = j | m^{(k)}(t-1) = i)$.

—**Normalize this transition matrix**:
This step is introduced by Gaveau and Shulman [9], [10], [11].
Let $\tau^{(k)}$ be a stochastic matrix with 2^k dimensions obtained after the k^{th} iteration. To assume that $\tau^{(k)}$ is irreducible matrix so [10] it has only one eigenvalue of unit norm to be called λ_0, and order the other eigenvalues by decreasing modulus.
λ_0 is associated with a strictly positive eigenvector; the remaining $2^k - 1$ eigenvectors satisfy $\lambda_0 \equiv 1 > |\lambda_1| \geq |\lambda_2| \dots$, correspond to right and left eigenvectors that are, respectively, v_m and l_m, with $(m \in 2^k)$regions, so:

$$\tau^{(k)} v_m = \lambda_m v_m \qquad l_m \tau^{(k)} = \lambda_m l_m$$

l_m was normalized by $\langle l_m | v_b \rangle = \delta_{mb}$, and v_0 is naturally normalized by $\sum_i v_0(i) = 1$.

—**Select an alphabet size of symbolic sequence**
The choice of the alphabet size plays a crucial role in symbolic time series analysis.
For example, a small value of k iteration may prove inadequate for capturing the characteristics of the system dynamic. On the other hand, a large value may lead to redundancy and waste of computational resources. The selection of optimal is an area of active research.
An entropy rate approach has been adopted for selecting the alphabet size [23].
Let $h(M^{(k)})$ denote the entropy rate of the transition matrix for iteration k[4].

$$h(M^{(k)}) = -\sum_{i,j} \pi_i^{(k)} \tau_{i,j}^{(k)} \log \tau_{i,j}^{(k)} \tag{13}$$

where the $\pi^{(k)}$ and $\tau^{(k)}$ are both maximum likelihood estimate of, respectively, the stationary distribution and probability of transition and log is taken as natural logarithm.
The value of entropy rate $h(M^{(k)})$ increases of iteration with k. We choose $k^* \in \{1, k_{max}\}$ where $h(M^{(k)})$ is maximal (practically $k_{max} \approx 10$).
To further increment $k > k^*$ value, leads to the problem of finite size of symbolic sequence and then the value of entropy rate decreases. This decrease is due to the finite size T of the symbolic sequence.
This effect was shown in our previous work [13] by partionning several time series of Lorenz system with different size T using subdivision method.
Also this effect is shown analytically in calculating the entropy rate of a given symbolic sequence $M^{(k)}$ of size T obtained by using the subdivision partition and the assumption of uniform transition probabilities. Shannon's entropy $H(M^{(k)})$ is written as:

$$\begin{aligned} H(M^{(k)}) &= -\sum_{m \in M^{(k)}} \tau^{(k)}(m) \log \tau^{(k)}(m) \\ &= -2^k \frac{T}{2^k} \log\left(\frac{T}{2^k}\right) \\ &= -T \log\left(\frac{T}{2^k}\right) \\ &= -T(\log T - k \log 2) \end{aligned} \tag{14}$$

And the entropy rate of $M^{(k)}$ is:

$$\begin{aligned} h(M^{(k)}) &= \lim_{k \to \infty} \frac{H(M^{(k)})}{k} \\ &= \lim_{k \to \infty} \frac{T}{k}(k \log 2 - \log T) \\ &= \lim_{k \to \infty} \left(T \log 2 - \frac{T \log T}{k}\right) \end{aligned} \tag{15}$$

In addition, for a sequence of finite size T obtained by the partition of subdivision, there will be a maximum of k iterations such as $T = 2^k$ and $k \to (\frac{\log T}{\log 2})$ and no ∞, so (15) equation can be written as:

$$h(M^{(k)}) = \lim_{k \to (\frac{\log T}{\log 2})} \left(T \log 2 - \frac{T \log T}{k}\right) = 0 \tag{16}$$

And the rate of entropy converges to zero for $k >> k^*$.

Now, different partitions process will be tested with various dynamical system in order to extract the best Markov partition.

3. RESULTS

Numerical simulations were used to check the validity of the approach on systems with known properties. Multivariate independent and identically distributed Gaussian process and deterministic chaotic system were used to compare these different partition methods.

3.1 White noise

In this section two samples of white noise will be used, the first is a Gaussian noise following a normal distribution with given mean and variance and the second is a uniform noise. Two random systems of 5 dimensions with $T = 30000$ points were generated.
Presume that the white noise is a realization of a random process in which the power spectral density is the same for all frequencies, we can deduce that the theoretical value of the entropy rate is equal to $\log(2^k)$ with 2^k is the number of symbols of the sequence obtained after each iteration k.
The estimation of the entropy rate of the transition matrix calculated from each sequence obtained is depicted in figure 1.
On this figure, the partitions of k-means, the subdivision and the varcluster give correct values of the rate of entropy coinciding with the theoretical values estimated by the blue line in both cases, but the partition of the SOM gives values below the theoretical line. On the other hand, the effect of the finite size of the sequence is shown for $k > k^*$ when $k^* = 5$.

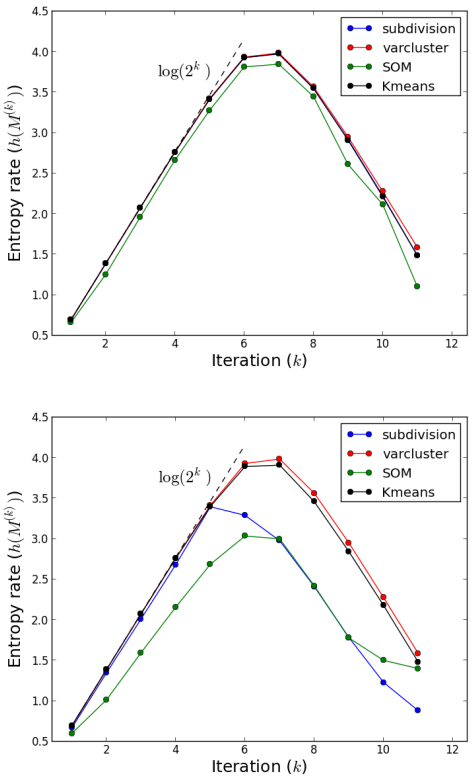

Fig. 1. Entropy rate of the meso-scale Markov process $(h(M^{(k)}))$ of white noise at each iteration (k) calculated in different partitions method. Top: Noise uniform. Bottom: Gaussian noise.

In the next section, these discretization methods will be applied on the phase space of a chaotic system (Lorenz attractor).

3.2 Lorenz chaotic system

The Lorenz system is a three dimensional dynamic system (topology and metric values are well studied in [19], [17], [18], [2] et [3]) defined by three nonlinear differential equations:

$$\begin{aligned} \dot{x} &= \sigma(y - x) \\ \dot{y} &= Rx - y - xz \\ \dot{z} &= -bz + xy \end{aligned} \tag{17}$$

It is based on three control parameters: the number of Pradtl σ, a ratio R and a parameter b dimensional rolls. (For more details see [17]).
The asymptotic motion in the phase space of this system is related to a chaotic attractor representing an axial symmetry, which is adjusted by control parameters such as $\sigma = 10$, $R = 28$ and $b = 8/3$.
Figure 2 depicts 16 regions of Lorenz attractor obtained after four steps of the subdivision procedure.

Figure 3 includes the values of the entropy rate calculated after a refinement of the partition of phase space.
This figure shows the importance of this refinement that correctly describe the dynamics following a convergence of the theoretical value of the entropy rate (≈ 0.92 for the Lorenz attractor computed

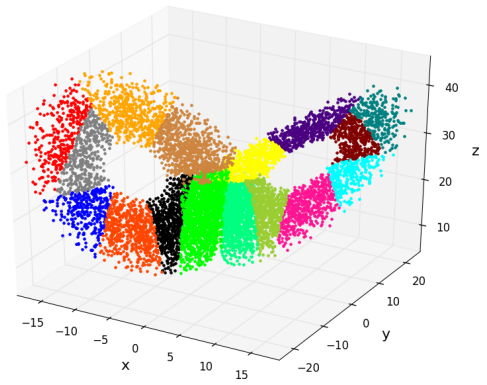

Fig. 2. Lorenz attractor after four steps of the subdivision procedure. Each color corresponds to one of the 16 meso-states of the system defined by Ulam subdivision procedure.

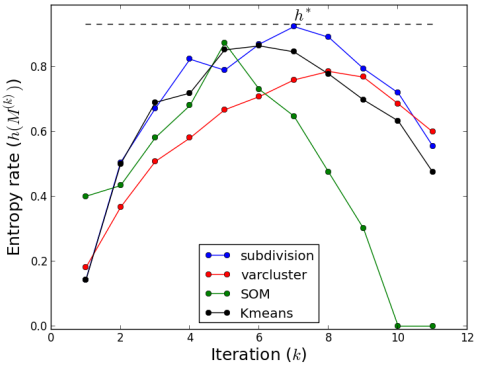

Fig. 3. Entropy rate of the meso-scale Markov process $(h(M^{(k)}))$ of Lorenz attractor at each iteration (k) calculated in different partitions method.

with a natural logarithm).
By subdivision process, it was able to reach this theoretical value after $k^* = 7$ iterations and to obtain a discrete Markov process with 128 states.
As for the case of multivariate random process, an observation of the finite-size effects appears for $k > k^*$.

4. DISCUSSION

Discretization of phase space of a dynamic system requires choosing the correct partition that preserves the dynamic properties of the system.
In this work, four partitions methods were compared among the most used such as: k-means, SOM, varcluster and subdivision.
This work is a continuation or even a step to verify a discretization in our previous paper [13] made to transform the microscopic space of measuring brain activity in a mesoscopic space.
Indeed, during this task we compared only two different methods of partition such as subdivision and varcluster and we concluded that the subdivision allowed to obtain satisfactory results to partitioning chaotic and stochastic systems.
We check here whether the methods of vector quantification can give better results. Thus we have used in this work, two other meth-

ods of vector quantification such as k-means and SOM. These four methods were applied to partition the same systems. We concluded that the subdivision method indeed gives the best results.
In fact, this comparison was based on the estimation of the entropy rate of transition matrix of Markov chains. Adding another parameter of comparison is one of the perspectives of this work.
It should also be noted that the application to other dynamic systems of these partition methods is needed to further verify that the subdivision process is among the best ways to discretize the phase space of dynamical systems.

5. REFERENCES

[1] C. Allefeld, H. Atmanspacher, and J. Wackermann. Mental states as macrostates emerging from brain electrical dynamics. *Chaos*, 19:015102, 2009.

[2] Erick.M Bollt. Review of chaos communication by feedback control of symbolic dynamics. *Bifurcation and Chaos*, 13:269–285, 2003.

[3] J.P. Bouchaud and P. Doussal. Numerical study of a d-dimensional periodic Lorentz gas with universal properties. *Statistical Physics*, 41:225–248, 1985.

[4] G. Ciuperca and V. Girardin. On the estimation of the entropy rate of finite Markov chains. *Proceedings of the International Symposium on Applied Stochastic Models and Data Analysis*, 2005.

[5] J. P. Crutchfield and N. H. Packard. Symbolic dynamics of noisy chaos. *Physica D*, 7:201–223, 1983.

[6] Pelleg Dan and Moore Andrew. Accelerating exact K-means algorithms with geometric reasoning. In Surajit Chaudhuri and David Madigan, editors, *Proceedings of the Fifth International Conference on Knowledge Discovery in Databases*, pages 277–281. AAAI Press, aug 1999.

[7] M. Dellnitz and A. Hohmann. A subdivision algorithm for the computation of unstable manifolds and global attractors. *Numerische Mathematik*, 75:293–317, 1997.

[8] Michael Dellnitz and Oliver Junge. *Set oriented numerical methods for dynamical systems*, volume 2. Elsevier, 2002.

[9] B. Gaveau and L.S. Shulman. Theory of nonequilibrium first-order phase transitions for stochastic dynamics. *Mathematical Physics*, 39:1517–1533, 1997.

[10] B. Gaveau and L.S. Shulman. Dynamical distance: coarse grains, pattern recognition, and network analysis. *Sciences Mathematiques*, 129:631–642, 2005.

[11] B. Gaveau, L.S. Shulman, and L.J. Shulman. Imaging geometry through dynamics: the observable representation. *Physics A*, 39:10307–10321, 2006.

[12] J Hadamard. Les surfaces à courbures opposées et leurs lignes géodésiques. *Biological Cybernetics*, 43:59–69, 1898.

[13] A. Hadriche, L. Pezard, J.P. Nandrino, H. Ghariani, A. Kachouri, and K.V. Jirsa. Mapping the dynamic repertoire of the resting brain. *NeuroImage Journal*, 78:448–62, September 2013.

[14] J. A. Hartigan and M. A. Wong. A k-means clustering algorithm. *Applied Statistics*, 28:100–108, 1979.

[15] H. Keller and G. Ochs. *Numerical approximation of random attractors*. Springer, 1999.

[16] T. Kohonen. Self-organized formation of topologically correct feature maps. *Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability*, 1:281–297, 1982.

[17] Christophe Letellier. *Caractérisation topologique et reconstruction d'attracteurs étranges*. PhD thesis, Université de Paris, 1994.

[18] Christophe Letellier. *systèmes dynamiques complexes: de la caractérisation topologique à la modélisation*. PhD thesis, 1998.

[19] E.N. Lorenz. Deterministic nonperiodic flow. *The Atmospheric Sciences*, 20:130–142, 1963.

[20] J.B MacQueen. Some methods for classification and analysis of multivariate observations. *Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability*, 1:281–297, 1967.

[21] Mohd Shukor, Zamzarina, Md. Sap, and Mohd. Noor. Clustering technique in data mining: general and research perspective. *Teknologi Maklumat*, 14:50–63, 2002.

[22] S. Ulam. *A collection of mathematical problems.* Interscience Publishers, New York, 1960.

[23] Rajagopalan. Venkatesh, Ray. Asok, Samsi. Rohan, and Mayer. Jeffrey. Pattern identification in dynamical sysytems via symbolic time series analysis. *Pattern recognition*, 40:97–07, 2007.