2011 International Symposium on Nonlinear Theory and its Applications
NOLTA2011, Kobe, Japan, September 4-7, 2011

Estimating a generating partition from distances

Yoshito Hirata and Kazuyuki Aihara

Institute of Industrial Science, The University of Tokyo,
4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan Email: yoshito@sat.t.u-tokyo.ac.jp

Abstract—Using a partition to convert a time series into a symbolic sequence is a common method to simplify the underlying dynamics. When the underlying dynamics is deterministic, there is a nice partition called generating, using which one can establish one-to-one correspondence between an orbit and an infinite symbolic sequence generated by the partition. In this talk, we extend the work of Hirata, Judd, and Kilminster, *Phys. Rev. E* **70**, 016215 (2004) so that one can estimate a generating partition given distances between a pair of times. By using artificial and real datasets, we demonstrate that the extended method can estimate generating partitions relatively accurately.

1. Introduction

Symbolic dynamics is a useful method to analyze time series data [1]. The symbolic dynamics helps to not only understand the underlying dynamics in a simple way [2] but also obtain topological and metric entropies with good accuracy [3, 4]. To construct the corresponding symbolic dynamics from a time series, one need to have a partition which divides the phase space into a finite number of regions. Among various partitions, generating partitions are the nicest partitions since they can establish one-to-one correspondence between a symbolic sequence so generated and an orbit. There are some approaches for estimating generating partitions for high-dimensional dynamics. The first proposed approach was to use homoclinic tangencies [5]. The second proposed approach was to use topological analysis [10, 11, 12]. The third approach was to assign symbolic substrings consistently to unstable periodic points [13, 14, 15]. As the fourth approach was to estimate generating partitions directly from time series [16, 17, 18].

If one can estimate a generating partition from distances between a pair of times, then one can construct the symbolic dynamics for various types of datasets such as time evolutions of point processes and networks. However, currently none of the above methods can be used to this setup.

In this paper, we extend the method of symbolic shadowing [17], a method for estimating a generating partition from a time series, to a method such that one can estimate a generating partition only from distances between every pair of times. The advantage of the proposed method is that we can drop the assumption that distances are the Euclidean.

The remaining parts of the paper are organized in the following way: In Section 2, we review symbolic shadowing [17]. In Section 3, we extend the symbolic shadowing so that one can estimate a generating partition from distances only. In Section 4, we demonstrate the extension by using examples. In Section 5, we conclude this paper.

2. Symbolic Shadowing

Let $\mathcal{M}$ be a metric space. Let $x(i) \in \mathcal{M}$ $(i = 1, 2, \ldots, n)$ be the i-th point of time series. Let $d : \mathcal{M} \times \mathcal{M} \to \{0\} \cup R^+$ be a distance function. For each pair (i, j) of times, we can obtain the corresponding distance $d(x(i), x(j))$. In this section, we consider the Euclidean norm. But, this point is extended in the next section.

Let $\mathcal{A}$ be a set of subsets $A_k \subset \mathcal{M}$ $(k = 0, 1, \ldots, S-1)$ where $A_k \cap A_l = \emptyset$ for $k \neq l$ and $\cup_{k=0}^{S-1} A_k = \mathcal{M}$. This $\mathcal{A}$ is called a partition. For each i, assign a symbol $s_i = k$ by $x(i) \in A_k$. Here S is the number of symbols as well as the number of the elements for the partition.

The substring for i with n_- backward and $(n_+ + 1)$ forward symbols is $s_{i-n_-} s_{i-n_-+1} \ldots s_{i-1}.s_i s_{i+1} \ldots s_{i+n_+}$ for $i = n_- + 1, \ldots, n - n_+$. Let us classify $\{x(i)\}$ using their substrings. To show a general substring with n_- backward and $(n_+ + 1)$ forward symbols, we use $\sigma_{-n_-} \sigma_{-n_-+1} \ldots \sigma_{-1}.\sigma_0 \sigma_1 \ldots \sigma_{n_+}$. Let $I(\sigma_{-n_-} \sigma_{-n_-+1} \ldots \sigma_{-1}.\sigma_0 \sigma_1 \ldots \sigma_{n_+})$ be a set of time indexes i such that $s_{i+j} = \sigma_j$ for $j = -n_-, -n_- + 1, \ldots, n_+$. Elements of such a set might be called members. Let $\Sigma_{[n_-, n_+]} : \mathcal{M} \to \{0, \ldots, S-1\}^{n_- + n_+ + 1}$ be a function such that $\Sigma_{[n_-, n_+]}(x(i)) = s_{i-n_-} \ldots s_{i-1}.s_i \ldots s_{i+n_+}$. In addition, for $\sigma_{-n_-} \sigma_{-n_-+1} \ldots \sigma_{-1}.\sigma_0 \sigma_1 \ldots \sigma_{n_+}$, choose a point in $\mathcal{M}$ called a representative, which is denoted by $r(\sigma_{-n_-} \sigma_{-n_-+1} \ldots \sigma_{-1}.\sigma_0 \sigma_1 \ldots \sigma_{n_+})$.

It is the result of Ref. [17] that we can choose a set of representatives such that $\sup_{x \in M} d(x, r(\Sigma_{[n_-, n_+]}(x))) \to 0$ as $n_-, n_+ \to \infty$ if and only if the partition is generating.

Practically, a generating partition can be estimated using symbolic shadowing [17]. In this method, the following optimization problem is solved approximately:

$$\min_{\{s_i\}, \{r(\cdot)\}} \sum_{i=n_+ + 1}^{n-n_-} \|x(i) - r(\Sigma_{[n_-, n_+]}(x))\|^2. \qquad (1)$$

The cost function is minimized as similarly as the expectation-maximization algorithm. Namely, we first fix a symbolic sequence and minimize the cost function over the representatives. Then, we fix the representatives and minimize the cost function over the symbolic sequence. And we repeat this iterations until it converges. The details of the algorithm are as follows:

(1): Generate a random symbolic sequence whose length is the same as the given time series.

(2): Let a be the initial length of substrings. Let $n_- = \lfloor a/2 \rfloor$ and $n_+ = \lfloor (a-1)/2 \rfloor$.

(3): For each point $x(i)$, find the closest representative $r(\sigma_{-n_-}\sigma_{-n_-+1}\dots\sigma_{-1}.\sigma_0\sigma_1\dots\sigma_{n_+})$ and assign $s(i)$ as $s(i) = \sigma_0$.

(4): For each existing substring $\sigma_{-n_-}\dots\sigma_{-1}.\sigma_0\dots\sigma_{n_+}$, find its representative by minimizing $\sum_{i=n_++1}^{n-n_-}\|x(i) - r(\Sigma_{[n_-,n_+]}(x(i)))\|^2$. Namely, find the representative by

$$
\begin{aligned}
& r(\sigma_{-n_-}\sigma_{-n_-+1}\dots\sigma_{-1}.\sigma_0\sigma_1\dots\sigma_{n_+}) \\
= & \frac{\sum_{i\in I(\sigma_{-n_-}\sigma_{-n_-+1}\dots\sigma_{-1}.\sigma_0\sigma_1\dots\sigma_{n_+})} x(i)}{|I(\sigma_{-n_-}\sigma_{-n_-+1}\dots\sigma_{-1}.\sigma_0\sigma_1\dots\sigma_{n_+})|}.
\end{aligned} \tag{2}
$$

(5): If Steps (3) and (4) do not change the symbolic sequence and the set of representatives, go to Step (6). Otherwise, go to Step (3).

(6): If the length of substring reached the pre-assigned maximum length, then finish the algorithm. Otherwise, increase a by $a \leftarrow a+1$, increment n_- and n_+ by $n_- = \lfloor a/2 \rfloor$ and $n_+ = \lfloor (a-1)/2 \rfloor$, and go back to Step (3).

3. An Extension

In this section, we extend the symbolic shadowing [17] so that we can estimate a generating partition from distances between pairs of times only. To realize this extension, we need to eliminate the averaging operation appearing in Step (4) of the above algorithm. Therefore, we replace the averaging by finding the member for each substring that has the minimum average distance between the members.

The details of the algorithm are as follows:

(1): Generate a random symbolic sequence whose length is the same as the given time series.

(2): Let a be the initial length of substrings. Let $n_- = \lfloor a/2 \rfloor$ and $n_+ = \lfloor (a-1)/2 \rfloor$.

(3): For each point $x(i)$, find the closest representative $r(\sigma_{-n_-}\sigma_{-n_-+1}\dots\sigma_{-1}.\sigma_0\sigma_1\dots\sigma_{n_+})$ and assign $s(i)$ as $s(i) = \sigma_0$.

(4): For each existing substring $\sigma_{-n_-}\dots\sigma_{-1}.\sigma_0\dots\sigma_{n_+}$, find its representative by finding the representative as

$$
\begin{aligned}
& r(\sigma_{-n_-}\sigma_{-n_-+1}\dots\sigma_{-1}.\sigma_0\sigma_1\dots\sigma_{n_+}) = \\
& \underset{x(i):i\in I(\sigma_{-n_-}\dots\sigma_{-1}.\sigma_0\dots\sigma_{n_+})}{\arg\min} \sum_{j\in I(\sigma_{-n_-}\dots\sigma_{-1}.\sigma_0\dots\sigma_{n_+})} d(x(i), x(j)).
\end{aligned} \tag{3}
$$

(5): If Steps (3) and (4) do not change the symbolic sequence and the set of representatives, go to Step (6). Otherwise, go to Step (3).

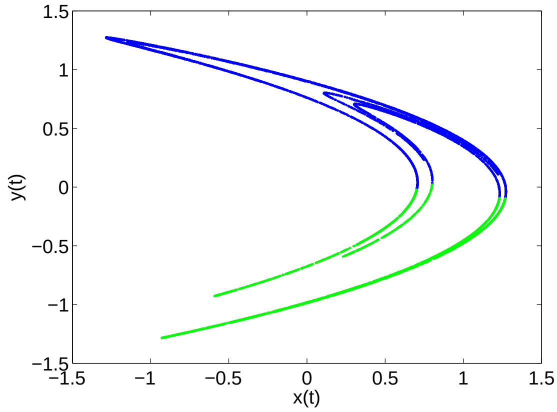

Figure 1: The estimated partition for a time series generated from the Hénon map.

(6): If the length of substring reached the pre-assigned maximum length, then finish the algorithm. Otherwise, increase a by $a \leftarrow a+1$, increment n_- and n_+ by $n_- = \lfloor a/2 \rfloor$ and $n_+ = \lfloor (a-1)/2 \rfloor$, and go back to Step (3).

Please confirm that the above algorithm works if we can obtain the distance for every pair of time points. Therefore, we can drop the assumption that the used distances are Euclidean. We call this algorithm the extended symbolic shadowing.

4. Examples

4.1. Hénon map

First, we applied the extended symbolic shadowing to a time series generated from the Hénon map. Here, the length of time series was 10 000. Then, we generated a random binary symbolic sequence of length 10 000. We started the extended symbolic shadowing with the substrings of length 2 and stopped it when the substrings reached the length of 12 and the algorithm converged. The obtained partition (Fig. 1) looked similar to the ones in Refs. [5, 17]. Out of 10 000 symbols, there were 12 different symbols between the result of the symbolic shadowing and that of its extension.

4.2. Ikeda map

Second, we applied the extended symbolic shadowing to a time series generated from the Ikeda map. The conditions were the same as the case of the Hénon map. The obtained partition (Fig. 2) looked similar to the ones in Refs. [13, 17]. There were 19 different symbols out of 10 000 symbols between the result of the symbolic shadowing and the one of its extension.

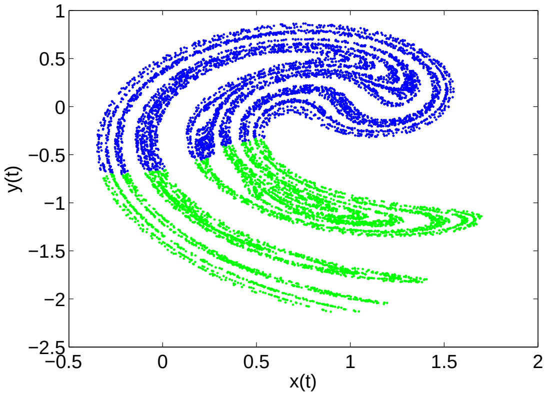

Figure 2: The estimated partition for a time series generated from the Ikeda map.

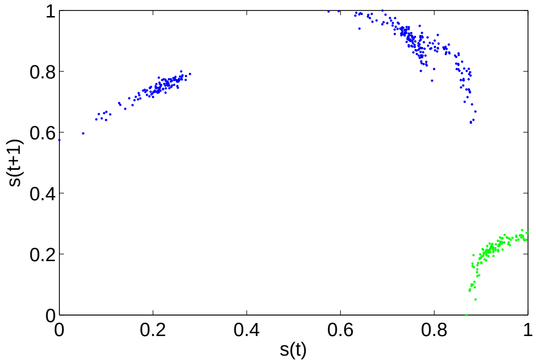

Figure 3: The estimated partition for a time series of the squid giant axon [19].

4.3. Squid giant axon

In addition, we applied the extended symbolic shadowing to the time series of the giant squid axon [19]. Here, we discarded the first 100 points to eliminate the transient part of the dataset. Then, we started the algorithm with a random binary symbolic sequence using the substrings of length 2, and stopped it when the length of substrings reached 8 and the algorithm converged. The obtained partition (Fig. 3) was exactly the same as the one obtained in Ref. [2].

These three examples showed that the extended symbolic shadowing worked well when generating partitions exist.

5. Conclusions

In this paper, we have extended the method for symbolic shadowing so that one can estimate a generating partition only from a set of distances between every pair of times. We demonstrated that the extension works well using time series of the Hénon map, the Ikeda map, and the squid giant axon. The proposed extension will provide insight into the complex dynamics described by time evolutions of point processes and networks.

Acknowledgments

This work is supported by the Aihara Innovative Mathematical Modelling Project, the Japan Society for the Promotion of Science (JSPS) through its “Funding Program for World-Leading Innovative R&D on Science and Technology (FIRST Program),” initiated by the Council for Science and Technology Policy (CSTP).

References

[1] C. S. Daw, C. E. A. Finney, and E. R. Tracy, “A review of symbolic analysis of experimental data,” *Rev. Sci. Instum.*, vol.74, pp.915–930, 2003.

[2] Y. Hirata, K. Judd, and K. Aihara, “Characterizing chaotic response of a squid axon through generating partitions,” *Phys. Lett. A*, vol.346, pp.141–147, 2005.

[3] M. B. Kennel and A. I. Mees, “Context tree modeling of observed symbolic dynamics,” *Phys. Rev. E*, vol.66, art.no.056209, 2002.

[4] Y. Hirata and A. I. Mees, “Estimating topological entropy via a symbolic data compression technique,” *Phys. Rev. E*, vol.67, art.no.026205, 2003.

[5] P. Grassberger and H. Kantz, “Generating partitions for the dissipative Hénon map,” *Phys. Lett.*, vol.113A, pp.235–238, 1985.

[6] F. Giovannini and A. Politi, “Homoclinic tangencies, generating partitions and curvature of invariant manifolds,” *J. Phys. A: Math. Gen.*, vol.24, pp.1837–1848, 1991.

[7] F. Christiansen and A. Politi, “Generating partition for the standard map,” *Phys. Rev. E*, vol.51, pp.R3811–R3814, 1995.

[8] F. Christiansen and A. Politi, “Symbolic encoding in symplectic maps,” *Nonlinearity*, vol.9, pp.1623–1640, 1996.

[9] L. Jaeger and H. Kantz, “Structure of generating partitions for two-dimensional maps,” *J. Phys. A: Math. Gen.*, vol.30, pp.L567–L576, 1997.

[10] M. Lefranc, P. Glorieux, F. Papoff, F. Molesti, and E. Arimondo, “Combining topological analysis and symbolic dynamics to describe a strange attractor and its crisis,” *Phys. Rev. Lett.*, vol.73, pp.1364–1367, 1994.

[11] G. Boulant, S. Bielawski, D. Derozier, and M. Lefranc, "Experimental observation of a chaotic attractor with a reverse horseshoe topological structure," *Phys. Rev. E*, vol.55, pp.R3801–R3804, 1997.

[12] G. Boulant, M. Lefranc, S. Bielawski, and D. Derozier, "Horseshoe templates with global torsion in a driven laser," *Phys. Rev. E*, vol.55, pp.5082–5091, 1997.

[13] R. L. Davidchack, Y.-C. Lai, E. M. Bollt, and M. Dhamala, "Estimating generating partitions of chaotic systems by unstable periodic orbits," *Phys. Rev. E*, vol.61, pp.1353–1356, 2000.

[14] J. Plumecoq and M. Lefranc, "From template analysis to generating partitions I: Periodic orbits, knots and symbolic encodings," *Physica D*, vol.144, pp.231–258, 2000.

[15] J. Plumecoq and M. Lefranc, "From template analysis to generating partitions II: Characterization of the symbolic encodings," *Physica D*, vol.144, pp.259–278, 2000.

[16] M. B. Kennel and M. Buhl, "Estimating good discrete partitions from observed data: symbolic nearest neighbors," *Phys. Rev. Lett.*, vol.91, art.no.084102, 2003.

[17] Y. Hirata, K. Judd, and D. Kilminster, "Estimating a generating partition from observed time series: Symbolic shadowing," *Phys. Rev. E*, vol.70, art.no.016215, 2004.

[18] M. Buhl and M. B. Kennel, "Statistically relaxing to generating partitions for observed time-series data," *Phys. Rev. E*, vol.71, art.no.046213, 2005.

[19] A. Mees, K. Aihara, M. Aadachi, K. Judd, T. Ikeguchi, and G. Matsumoto, "Deterministic prediction and chaos in squid axon response," *Phys. Lett. A*, vol.169, pp.41–45, 1992.