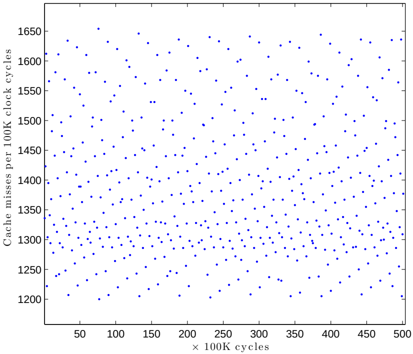
Topological entropy as a measure of computer system performance

Zach Alexander, Todd Mytkowicz, Liz Bradley

University of Colorado at Boulder

December 5, 2009

To measure performance, we record the number of cache misses per 100K clock cycles for various combinations of software and hardware.



The trace shown here from a program that initializes a matrix in row-major order on an Intel Core2 processor.[4]

What type of questions should we be asking?

- Is the data random? deterministic? periodic?
- Can we extract information about the underlying dynamics of the program, the computer, or the interaction of the two?

Big Question:

Is it worthwhile to use/develop topological tools to answer these questions?

Symbolic Dynamics

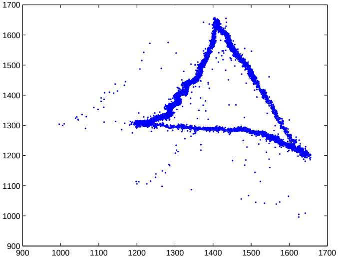
Symbolic dynamics is an approach that can be used to analyze complex data sets. A symbol sequence is created by partitioning the data and identifying each point with the element of the partition that it belongs to.

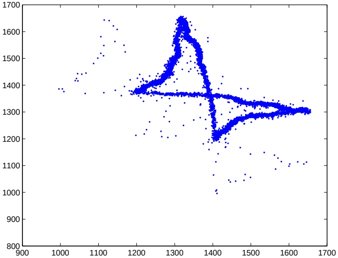
In practice, the experimental data is divided by a partition with as few as two elements. This creates a symbol sequence on two symbols.

e.g. 01011010001011...

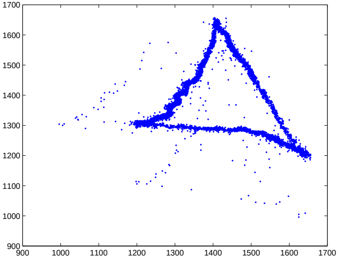
Advantage:

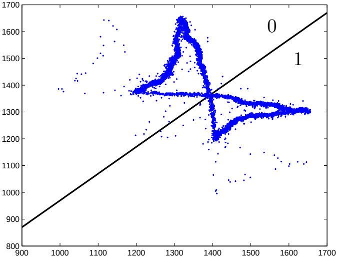
Very robust for noisy data sets





Return maps are used to generate a partition. First and second return maps are shown here.

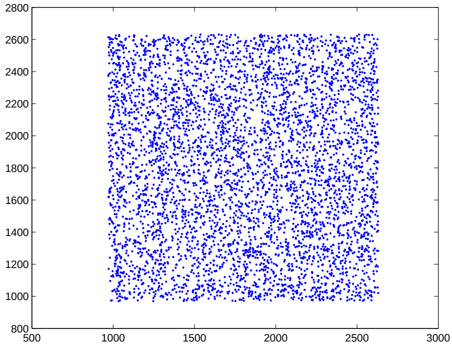




Return maps are used to generate a partition. First and second return maps are shown here.

A return map of random data

For contrast, this is a typical first return map for randomly generated data with similar characteristics to row_maj



Shannon Entropy

The Shannon entropy of a data set $X = \{x_1, \ldots, x_n\}$ is defined as

$$H(X) = -\sum_{i=1}^{n} p(x_i) \log_b p(x_i).$$

Roughly, this quantity measures the information content of a data set. For example, $H(\text{row\_maj}) = .9924$. That is, the symbols 0 and 1 are equally likely.

Topological Entropy

Shannon entropy gives information about a sequence of *random* events. A snippet of the symbol sequence generated from row_maj indicates that this may not be a good assumption.

$$\ldots 0011001100110001100110 01\ldots$$

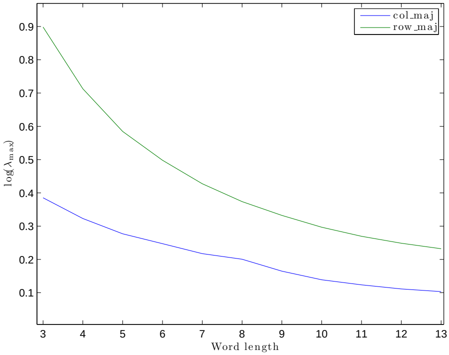
Topological entropy is a measure of the complexity of a deterministic dynamical system. It is defined as

$$h_T(X) = \lim_{n\to\infty} \frac{\log N(n)}{n}$$

where $N(n)$ is the number of words of length n occurring in the symbol sequence.

Topological entropy from a transition matrix

An upper bound for the topological entropy of a dynamical system can be calculated from a transition matrix.

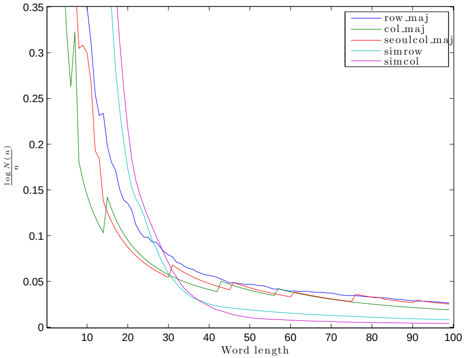


Example

The transition matrix $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ Corresponds to a topological entropy of .4812.

Topological entropy from counting word occurrences

A lower bound for the topological entropy is computed from the formula given in the definition.

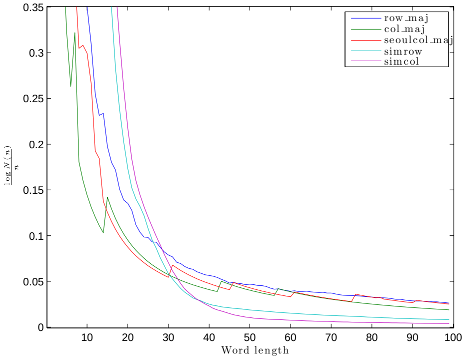


Example

The sequence 0011001100110001100110011001 contains five words of length 3.

Topological entropy from counting word occurrences

A lower bound for the topological entropy is computed from the formula given in the definition.

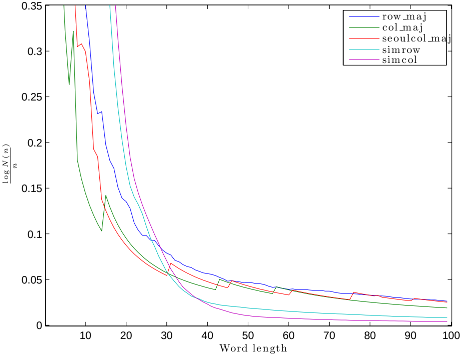


Example

The sequence 0011001100110001100110001 contains five words of length 3.

Topological entropy from counting word occurrences

A lower bound for the topological entropy is computed from the formula given in the definition.

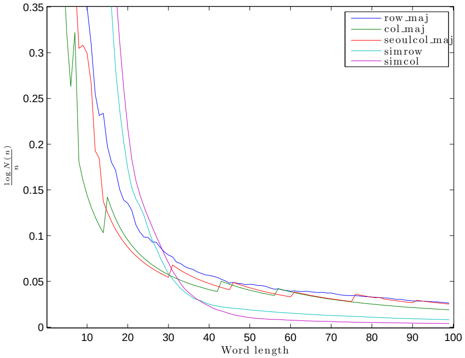


Example

The sequence 001100110011000110011001 contains five words of length 3.

Topological entropy from counting word occurrences

A lower bound for the topological entropy is computed from the formula given in the definition.

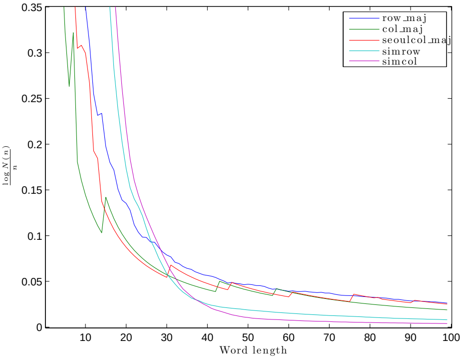


Example

The sequence 00110011001100011001 1001 contains five words of length 3.

Topological entropy from counting word occurrences

A lower bound for the topological entropy is computed from the formula given in the definition.

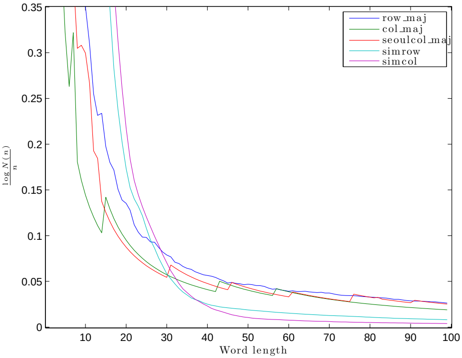


Example

The sequence 00110011001100011001100 1 contains five words of length 3.

Topological entropy from counting word occurrences

A lower bound for the topological entropy is computed from the formula given in the definition.

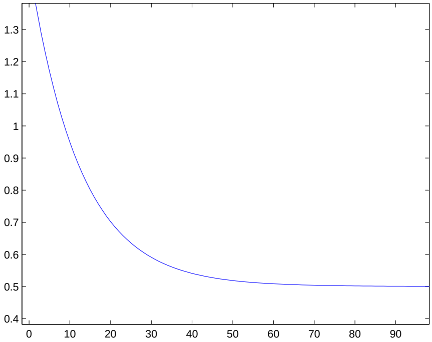


Example

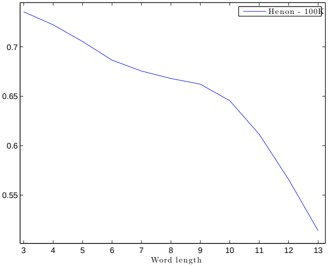
The sequence 0011001100110001100110001 contains five words of length 3.

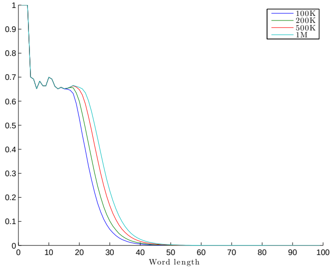
Data Length

Ideally, we want the entropy curves to completely flatten out. The fact that they do not is a factor of the length of the data set.



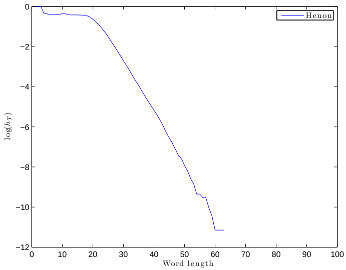
Topological entropy of the Henon map

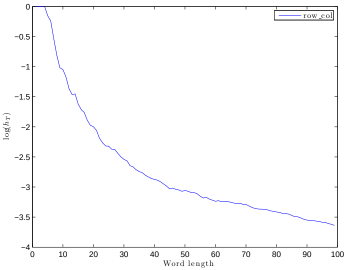




Here we can see the effects of data length on our calculations. The Henon map is a well studied example.[1]

Logarithmic decay





The logarithms of the entropy functions suggest a meaningful wordlength.

Conclusion

Cons

- Both of the algorithms discussed require extremely long data sets and considerable computational power.
- As shown by the example of the Henon map, a large increase in data length gives a minimal increase in accuracy

Pros

- Consistent differences between row_maj and col_maj on a real machine, and between the real machines and the simulator using the transition matrix algorithm and the word length algorithm, respectively, suggests that topological entropy contains information about the physical system.
- We can determine that the cache misses are generated from a deterministic, non-periodic, dynamical system (chaos)

A note on *topology*

Topological entropy is so named because it is an invariant of the topology of the actual state space of the dynamical system. Although our computations were not precise enough to be conclusive, it does seem plausible that the actual topological entropies can differentiate between different computer systems. This suggests the use and development of more sophisticated measures of state space topology, e.g. [2], [3], and [5].

Big Answer:

Yes!!!

J Curry.
On computing the entropy of the henon attractor.
Journal of Statistical Physics, Jan 1981.

R Gilmore.
Topological analysis of chaotic dynamical systems.
Reviews of Modern Physics, 70(4):1455–1529, 1998.

Konstantin Mischaikow, M Mrozek, and J Reiss.
Construction of symbolic dynamics from experimental time series.
Physical review letters, Jan 1999.

T Mytkowicz, A Diwan, and E Bradley.
Computer systems are dynamical systems.
CHAOS: An Interdisciplinary Journal of Nonlinear Science, 19:033124, 2009.

Vanessa Robins.
Toward computing homology from finite approximations.
Topology Proceedings, Jan 1999.