Information Theoretic Aspects of Self-Organization

(Cosma R. Shalizi; Daniel Polani)

Stefan Klampfl
klampfl@igi.tugraz.at

Institute for Theoretical Computer Science (IGI)
Graz University of Technology, Austria

Computational Intelligence Seminar B, SS 2008
June 2, 2008
Introduction

Self-organization as increasing statistical complexity (Shalizi)

Self-organization via observers (Polani)

SC- vs O-self-organization

Discussion
Introduction

- Phenomenon of \textit{self-organization} is ubiquitous in the study of complex systems.
- As much agreement there is on whether self-organization is present or absent, as little agreement exists on the precise meaning of the word ("I know it when I see it", pragmatic characterizations).
- Formal foundations, precise mathematical definitions are missing.
- We want to quantify the amount of self-organization in a system.
- Self-organization becomes important for engineering.
Motivation

Why information theory?

- Many existing models have specific requirements to describe self-organization
  - differentiable manifolds, eigenvalues of Jacobian, etc.
  - Euclidean space
  - determinism
- It is easy to imagine scenarios where these requirements do not hold (e.g., social or food web networks)
- Information theory requires only minimal structure (a probability measure) on the space of interest
- Promising common language for the study of general systems
- Two approaches:
  - Self-organization as increasing statistical complexity
  - Self-organization via observers
Notation

- Consider a random variable $X$ assuming values $x \in \mathcal{X}$

- **Entropy of $X$:**

  $$H(X) := - \sum_{x \in \mathcal{X}} p(x) \log p(x)$$

- **Conditional entropy of $Y$ given $X$:**

  $$H(Y|X) := - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(y|x) \log p(y|x)$$

- The **joint entropy** of $X$ and $Y$ is the entropy of the random variable $(X, Y)$:

  $$H(X, Y) := - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log p(x, y)$$
The mutual information of $X$ and $Y$:

$$I(X; Y) := H(Y) - H(Y|X)$$

$$= H(X) + H(Y) - H(X, Y)$$

intrinsic information (multi-information, integration) for random variables $X_1, \ldots, X_k$:

$$I(X_1; \ldots; X_k) := \sum_{i=1}^{k} H(X_i) - H(X_1, \ldots, X_k)$$

measures the degree of dependence between the $X_i$. 
Self-organization as increasing statistical complexity

- based on $\epsilon$-machine formalism (Crutchfield and Young, 1989)
- consider a stochastic process $\mathbf{X}$ (infinite past and future)

$$\mathbf{X} = \ldots X(t-3)X(t-2)X(t-1)X(t)X(t+1)X(t+2)X(t+3)\ldots$$

\[ \xrightarrow{\sim} \mathbf{X} \text{ (past)} \quad \mathbf{X} \text{ (future)} \]

- equivalence relation $\sim_\epsilon$ on pasts $\xrightarrow{\sim}$

$$\xrightarrow{\sim} \sim_\epsilon \xrightarrow{\sim} \iff P(\xrightarrow{\sim} | \xrightarrow{\sim}) = P(\xrightarrow{\sim} | \xrightarrow{\sim}') \quad \forall \xrightarrow{\sim}$$

- partitions all possible pasts $\xrightarrow{\sim}$ into disjoint sets $\xrightarrow{\sim}$ ("causal states" of $\mathbf{X}$)
- defines a function $\epsilon : \xrightarrow{\sim} \rightarrow \xrightarrow{\sim}$
Self-organization as increasing statistical complexity

- $\epsilon$-machine of a process $X$:
  - deterministic automaton with states $\tilde{X}$
  - $t \rightarrow t + 1$: transition from $\tilde{x} = \epsilon(\tilde{x}(t))$ to $\tilde{x}' = \epsilon(\tilde{x}(t+1))$
  - transition is labelled with $x^{(t+1)}$

- $\epsilon$-machine is the unique minimal maximally predictive model of the original process (Shalizi and Crutchfield, 2001)

- *statistical complexity*:

  $$C_\mu(X) := H(\tilde{X})$$

  measure of the memory required to perform $X$

- A system can be called self-organizing if statistical complexity grows with time (Shalizi, 2001) (*SC-self-organization*)
Self-organization as increasing statistical complexity

Example: cyclic cellular automata (CCA) (Shalizi et al., 2004)

- Each site in a square lattice has one of $\kappa$ colors
- A cell of color $k$ will change to color $k + 1 \mod \kappa$ if there are at least $T$ cells of that color within distance $r$ ($\kappa = 4$, $r = 1$)
- (a) $T = 1$ incoherent local oscillations, (b) $T = 2$ spiral waves, (c) $T = 3$ meta-stable turbulence, (d) $T \geq 4$ fixation
- spatial extension of $\epsilon$-machine

(a) T = 1 incoherent local oscillations, (b) T = 2 spiral waves, (c) T = 3 meta-stable turbulence, (d) T \geq 4 fixation

Spatial extension of $\epsilon$-machine
Self-organization as increasing statistical complexity

- *excess entropy* is a lower bound on the statistical complexity

\[ E := I(\overrightarrow{X}; \overleftarrow{X}) \leq H(\tilde{X}) = C_\mu(X) \]

- \( E \) is the amount of *apparent* information in the observed behavior about the past
- \( C_\mu \) is the amount of memory that the process stores *internally* about its past

- Information is squeezed through the “bottleneck” given by the causal states \( \tilde{X} \) (Information Bottleneck principle, Tishby et al 1999)

\[
\min I(\overleftarrow{X}; \tilde{X}) - \beta I(\tilde{X}; \overrightarrow{X}) \quad \text{for } \beta \to \infty
\]
Self-organization via observers

- A *(perfect) observer* of a random variable $X$ is a collection $X_1, \ldots, X_k$ of random variables for which $H(X|X_1, \ldots, X_k) = 0$

- *organization information* with respect to the observer: $I(X_1; \ldots; X_k)$

- A system is called self-organizing (with respect to the given observer) if the organization information increases over time (Polani, 2003) *(O-self-organization)*

- Examples of observer variables:
  - partition of a system into subsystems
  - states of individual agents
  - points in space

- The whole system must have some degree of freedom, otherwise $I(X_1; \ldots; X_k) = 0$
Self-organization via observers

- Example: Kohonen’s Self-Organizing Map (SOM) (Polani, 2003)
  - set of units $i = 1, \ldots, k$ located on a grid, with weights $X_i \in \mathbb{R}^n$
  - trained with random samples $V \in \mathbb{R}^n$
  - weights $X_i$ reorganize such that neighboring features in $\mathbb{R}^n$ are represented by neighboring units
  - total state vector $S = (X_1, \ldots, X_k)$

- self-organization is immediately evident, but not obvious to quantify
Self-organization via observers

▶ analyze SOM with $4 \times 4$ units with weights $X_i \in [0, 1]^2$; 4 possible values for each weight component ($H(S^{(t=0)}) = 4 \times 4 \log_2 4^2 = 64$bit)

▶ information gain: $H(S^{(t=0)}) - H(S^{(t=t_{final})})$

▶ prediction entropy: $H(S^{(t+1)}|S^{(t)})$

▶ organization information: (weights $X_1, \ldots, X_k$ as observer)
$I(X_1; \ldots; X_k) = \sum_{i=1}^{k} H(X_i) - H(X_1, \ldots, X_k)$
Self-organization via observers

Consider 3 cases:

- begin with a random state $S^{(t=0)}$ and end with a random state $S^{(t=t_{\text{final}})}$ independent from $S^{(t=0)}$
  - information gain and organization information vanish
  - prediction entropy always yields the full entropy
  - all measures correctly identify the system as not self-organizing

- start with a random state and end up with a unique final state
  - information gain is $64\text{bit} - 0\text{bit} = 64\text{bit}$
  - organization information vanishes
  - “freezing” the system is not self-organizing

- original SOM dynamics
  - information gain is $64\text{bit} - 3\text{bit} = 61\text{bit}$
  - prediction entropy is unable to distinguish this from previous case
  - organization information increases

- organization information identifies just the organizational aspect of the dynamics
Self-organization via observers

- How does this measure change if one observer is replaced by another? (change of the “coordinate system”)
- Relation between a fine-grained observer $X_1, \ldots, X_k$ and a coarse-grained observer $\tilde{X}_1, \ldots, \tilde{X}_{\tilde{k}}$

$$
\begin{align*}
X_1, \ldots, X_k, X_{k_1+1}, \ldots, X_{k_2}, \ldots, X_{k_{\tilde{k}-1}+1}, \ldots, X_k \\
\tilde{X}_1 & \quad \tilde{X}_2 & \quad \ldots & \quad \tilde{X}_{\tilde{k}}
\end{align*}
$$

- Organization information of the fine-grained observer:

$$
I(X_1; \ldots; X_k) = I(\tilde{X}_1; \ldots; \tilde{X}_{\tilde{k}}) + \sum_{j=1}^{\tilde{k}} I(X_{k_j-1}+1; \ldots; X_{k_j})
$$

- “The fine-grained system is more than its coarse-grained version”
- Under recoding ($Y_i = f_i(X_1, \ldots, X_k)$) there is no canonical way of transforming organization information
SC- vs O-self-organization

- **SC-self-organization**
  - measures the growth of statistical complexity of a single time-series
  - concentrates on temporal dynamics
  - views self-organization as an intrinsic property of the system
  - specific spatial structure requires additional modifications

- **O-self-organization**
  - additionally requires a set of RV (observer) through which the system is observed
  - concentrates on compositional aspects of the system
  - measures self-organization with respect to this particular observer
  - spatial organization can be described easily

- similar in the general philosophy, but “orthogonal” as quantities

- which method to use depends on the system to analyze
Discussion

- Existing notions of self-organization are either too vague or too specific
- Information theory provides a precise language for identifying the conditions of self-organization
- It allows to quantify the amount of self-organization
- It can be applied to a wide range of systems
- Information-theoretic quantities can be hard to estimate
- First step towards a quantitative understanding of self-organization
Inferring statistical complexity.

Measuring self-organization via observers.

Foundations and formalizations of self-organization.
References II


*Causal architecture, complexity and self-organization in time series and cellular automata.*


Computational mechanics: Pattern and prediction, structure and simplicity.


Quantifying self-organization with optimal predictors.