Complexity seminar

Hugo Cisneros

March 10, 2020

Motivation

Measuring complexity

- Evolution of complexity can be seen as a basis for intelligent behavior.

Measuring complexity

- Evolution of complexity can be seen as a basis for intelligent behavior.
- Some cellular automata could be exhibiting evolutionary properties.

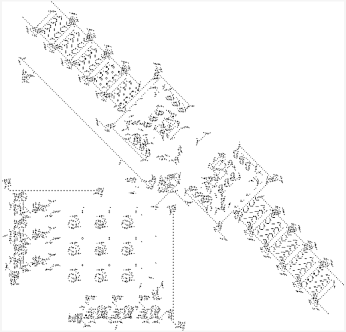


Figure 1: Turing machine in Game of Life

Measuring complexity

- Evolution of complexity can be seen as a basis for intelligent behavior.
- Some cellular automata could be exhibiting evolutionary properties.

But still many open questions:

- Evolution of complexity can be seen as a basis for intelligent behavior.
- Some cellular automata could be exhibiting evolutionary properties.

But still many open questions:

- How large is this subclass of cellular automata?

Measuring complexity

- Evolution of complexity can be seen as a basis for intelligent behavior.
- Some cellular automata could be exhibiting evolutionary properties.

But still many open questions:

- How large is this subclass of cellular automata?
- Can it happen faster in some systems (number of states, size of grids, dimensions)

Measuring complexity

- Evolution of complexity can be seen as a basis for intelligent behavior.
- Some cellular automata could be exhibiting evolutionary properties.

But still many open questions:

- How large is this subclass of cellular automata?
- Can it happen faster in some systems (number of states, size of grids, dimensions)

$\rightarrow$ **Measuring the complexity of these systems can help**

Complexity

Complexity and interestingness

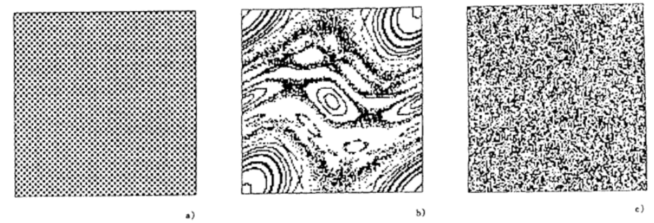


Figure 1: Three pictures with varying "complexities"

Some of Grassberger's features of a complex process (Grassberger 1989):

- Between disorder and order
- Often involves hierarchies with feedback loops
- Higher level concepts arise without being put explicitely

Definition

For a discrete random variable , with $p_i \triangleq P(X = i)$

$$H = -\sum_i p_i \log p_i$$

Largest when $p_1 = p_2 = ... = p_n = \frac{1}{n}$, uncertainty is maximal.

- Measures *randomness* of inputs
- High and low entropy correspond respectively to maximal order and disorder

Kolmogorov complexity

Definition

For a universal computer U the algorithmic information of S relative to U is defined as the length of the shortest program that yields S on U.

$$C_U(S) = \min_{Prog_U(S)} \text{Len}[Prog_U(S)]$$

- Theoretically close to what we are looking for
- Not computable $\rightarrow$ this makes it hard to use in practice

Measuring complexity with compression

Compression: Generalities

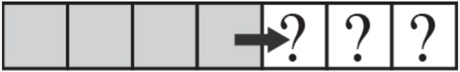
Goal: reduce the size of some data (not possible in general).

Essential in modern software: ZIP, PNG, JPEG, GIF, MP3, MP4, etc.

Some of those algorithms started as a measure of complexity (Lempel and Ziv 1976).

Prediction by partial matching compression

- Estimate the probability of the next symbol



- Encode optimally with respect to these predictions (with e.g arithmetic encoding)

Remark

Compression approximates Kolmogorov complexity because

```
compressed_string + decompressor_program
```

is a valid program that can generate the string.

The approximation works better with "intelligent" enough compressors.

A better compressor has a better “understanding” of data (Mahoney 1999; Zenil 2019).

Very compressible object have a simple underlying structure.

```
01001000 01100101 01101100
01101100 01101111 00100000
01110111 01101111 01110010
01101100 01100100 00100001
```

Compression and comprehension

A better compressor has a better "understanding" of data (Mahoney 1999; Zenil 2019).

Very compressible object have a simple underlying structure.

```
01001000 01100101 01101100
01101100 01101111 00100000
01110111 01101111 01110010
01101100 01100100 00100001
```

```
== Hello world!
```

Measuring complexity in cellular automata

Compression for complexity in CA

Compressing cellular automata in 1D (Zenil 2010):

- Treat the CA as a string

0101110

- Compress with common algorithms (Gzip)
- Use the length of the compressed string as the complexity metric

Compressing cellular automata in 1D (Zenil 2010):

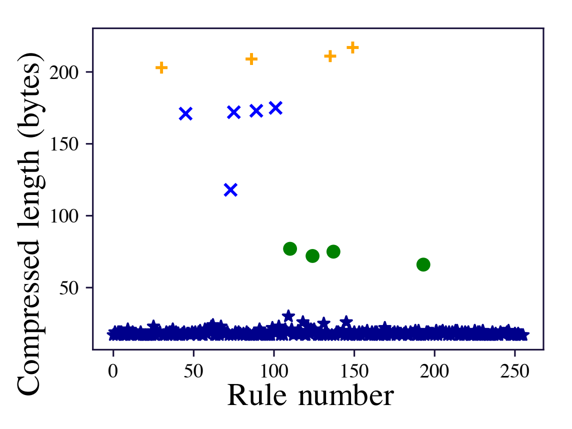


Figure 2: Compressed lengths of the 256 ECA

Compressing cellular automata in 1D (Zenil 2010):

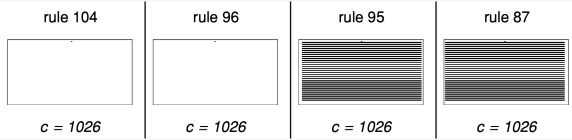


Figure 2: Lowest compressed length: Regular and periodic

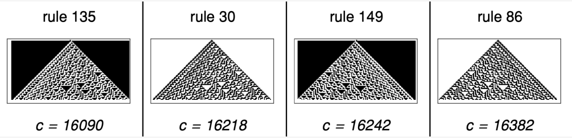
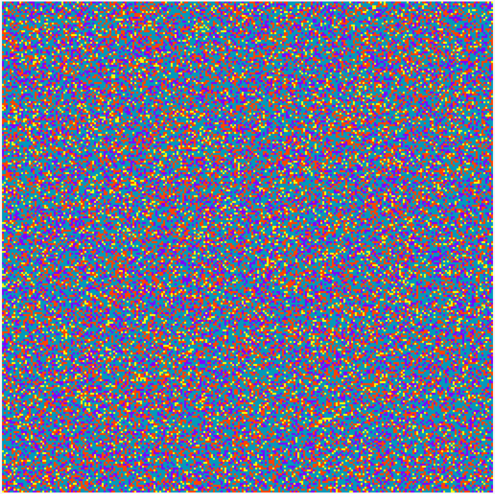


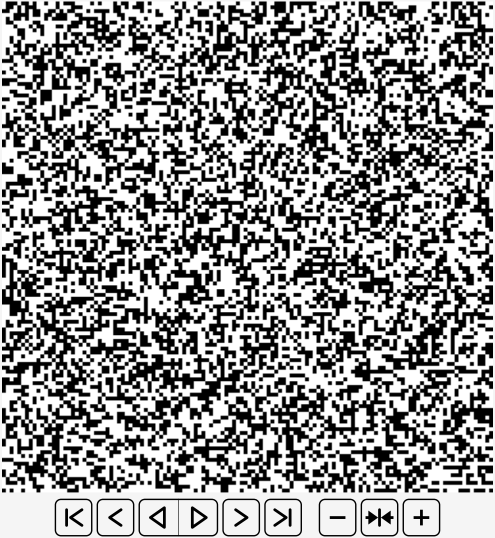
Figure 3: Highest compressed length: Disordered[1]

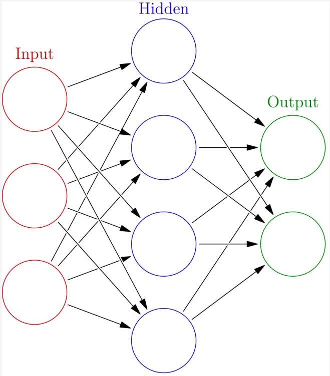
[1] Figures from Zenil 2010.

Results for compression as the only metric



Results for compression as the only metric





Input/output pairs $(x_i, y_i) \in \mathbb{R}^N \times \mathbb{R}^K$, $\mathbf{W}_1 \in \mathbb{R}^{N\times H}$, $\mathbf{W}_2 \in \mathbb{R}^{H\times K}$

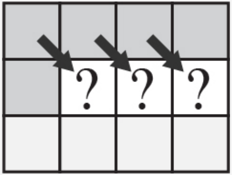
$$h_i = f_1(\mathbf{W}_1 x_i + \mathbf{b}_1)$$

$$\hat{y}_i = f_2(\mathbf{W}_2 h_i + \mathbf{b}_2)$$

Neural networks for compression

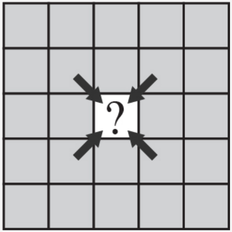
- Compressing text

- Compressing images



Neural networks for compression

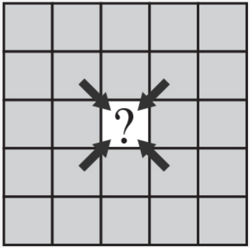
- Compressing arbitrary data with no decompression



We can use neural networks to make these predictions (Schmidhuber and Heil 1996; Mahoney 2000)

Neural networks for complexity

A neural network is trained on the input/output pairs like below.



The error of the network L quantifies how easily "learnable" the patterns are and how compressible the system is.

Compressing temporally

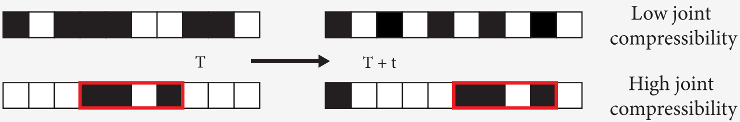
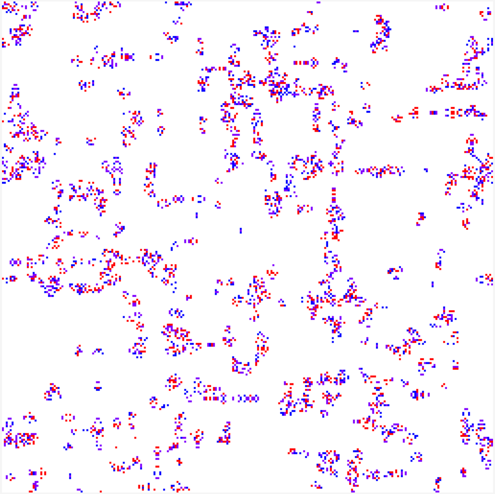


Figure 4: Compressing different timesteps together

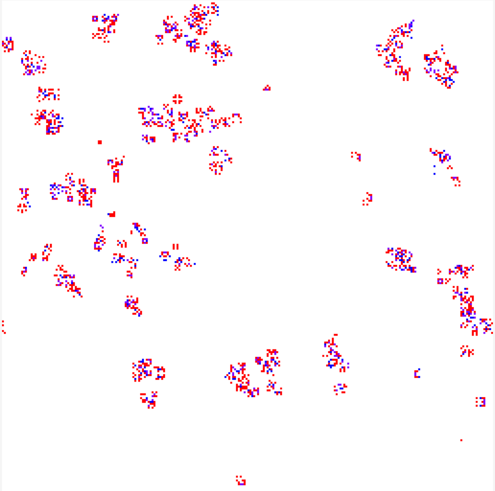
Training loss $L^{(T)}$ and testing losses $L^{(T+\tau)}$.

Score: $\frac{L^{(T)}}{L^{(T+\tau)}}$

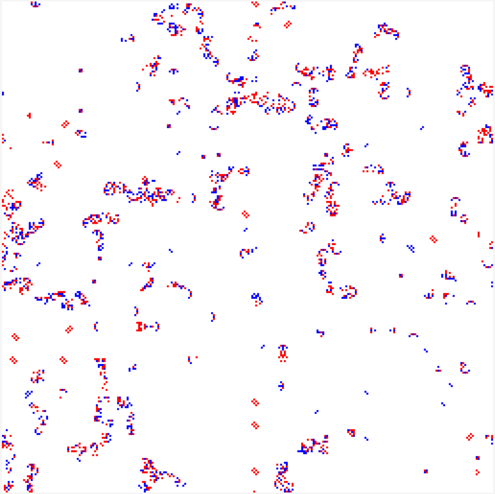
Interesting systems with high score



Interesting systems with high score



Interesting systems with high score



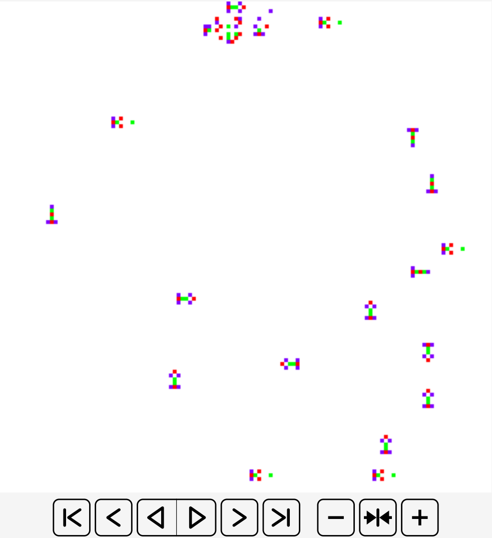
Next directions

Coarse-graining

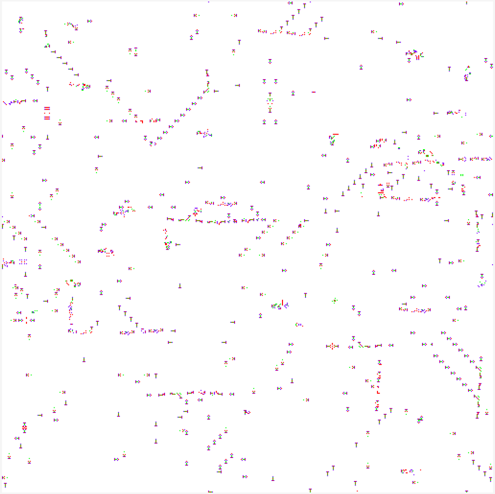
To detect larger structures and potentially more complex behavior, it might be necessary to

- Step back from the local approach
- Study very large grids

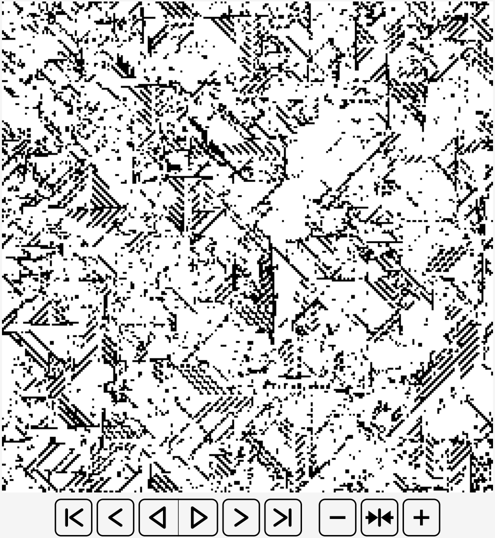
Why scale things up?



Why scale things up?

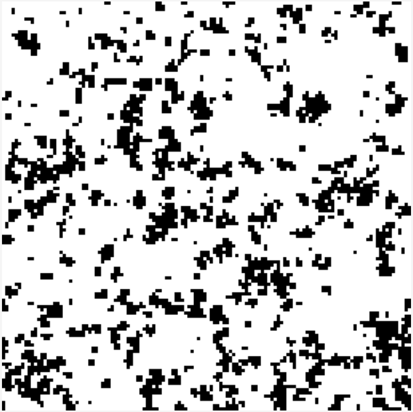


Why scale things up?



Idea: encode blocks of inputs according to how probable they are to appear.

- Very probable blocks are converted to a 0 pixel
- Improbable blocks are converted to a 1 pixel



Hourglass-shaped neural nework

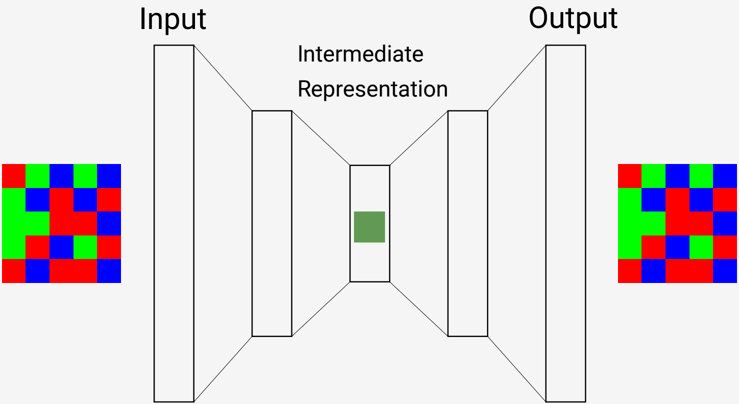


Figure 5: Autoencoders for coarse-graining

Remaining open questions

- Is this form of “interestingness” enough?
- Where should the search happen?
- Use/combine *theoretical* metrics with *observational* ones?
- Can complexity keep increasing in isolation?

Thank you!

References

Grassberger, Peter (1989). "Randomness, Information, and Complexity". In: *Proceedings of the 5th Mexican School on Statistical Physics*. Mexican School on Statistical Physics (EMFE). arXiv: 1208.3459.

Lempel, A. and J. Ziv (Jan. 1976). "On the Complexity of Finite Sequences". In: *IEEE Transactions on Information Theory* 22.1, pp. 75–81. ISSN: 0018-9448. DOI: 10.1109/TIT.1976.1055501.

Mahoney, Matthew V (1999). "Text Compression as a Test for Artificial Intelligence". In: *Proceedings of AAAI-1999*. AAAI, p. 3.

Mahoney, Matthew V (2000). “Fast Text Compression with Neural Networks”. In: *FLAIRS*. FLAIRS Conference, p. 5.

Schmidhuber, J. and S. Heil (Jan. 1996). “Sequential Neural Text Compression”. In: *IEEE Transactions on Neural Networks* 7.1, pp. 142–146. ISSN: 1045-9227. DOI: 10.1109/72.478398.

Zenil, Hector (2010). “Compression-Based Investigation of the Dynamical Properties of Cellular Automata and Other Systems”. In: *Complex Systems* 19.1.

– (Sept. 3, 2019). “Compression Is Comprehension, and the Unreasonable Effectiveness of Digital Computation in the Natural World”. In: arXiv: 1904.10258 [cs, math].