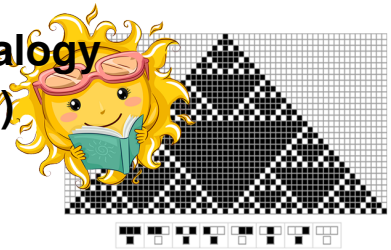
Information-Theoretic Analysis of ECA Rules (Genealogy Interceded Phenotypic Analysis (GIPA) of ECA rules)

R Goyal

August 24, 2022



This study demonstrates the grouping of the ECA Rules using information-theoretic measures in light of Wolfram's classification of ECA, with 88 rule equivalence classes. A formal qualitative behavior of the space-time diagrams for different rules gives rise to the macroscopic parameters, whose quantified (numerical) evaluation provides an opportunity for further analysis through machine learning-based approaches to extract interesting patterns. In this study, analysis has been done for a fixed single-cell scenario for different numbers of cells and iterations, with additional experiments for difference analysis with random inputs. This study is based on the prior studies by Borriello, and Chliamovitch. Adverting to the prior expositions, Hector Zenil has done pioneering work in algorithmic probability and analyzed information dynamics of cellular automata from different perspectives. This study, however, analyzes at the level of rules; therefore, we use the concept of BiEntropy, which was proposed by Grenville J. Croll to compute approximate information content of a binary string (see Eq. 1).

$$H(p) = \frac{1}{2^{n-1}-1}\sum_{k=0}^{n-2}(-p(k)\log_2 p(k) - (1-p(k))\log_2(1-p(k)))2^k \tag{1}$$

The amount of information processed by a CA rule in a space-time patch is captured through four measures.

- DiffEntropy (DE): Maximum absolute difference in the BiEntropy values of any reachable configuration (C_j) from the initial configuration (C_i), e.g.,

$$DE = Max(abs(BiEntropy(C_j) - BiEntropy(C_i))) \tag{2}$$

  This parameter captures the impact of the transformation carried out by a rule.

- SimConfigOrdered (SCO): Count the similar BiEntropy values (based on a threshold value, i,e, 0.01) of two configurations in a sorted list of entropy values.

$$SCO = BiEntropy(C_i) - BiEntropy(C_j) \tag{3}$$

  This parameter captures the frequency of similar (information content-wise) reachable configurations during processing.

- SimConfigImmediate (SCI): Count the similar BiEntropy values (based on a threshold value, i,e, 0.01) of two successive configurations.

$$SCI = BiEntropy(C_i) - BiEntropy(C_{i+1}) \tag{4}$$

- SimConfigFluctuation (SCF): Count the fluctuations in the BiEntropy values (where DE/2> 0.01) of two successive configurations.

$$SCF = abs(BiEntropy(C_i) - BiEntropy(C_{i+1})) > DE/2 \tag{5}$$

  This parameter captures the higher fluctuations (> DE/2) of successive reachable configurations during processing.

Category-I

- Entropy (BiEnrtropy values) does not change, though string may get change (like in Rule 51); because change still leads to a pattern (like 1's complement in the case of Rule 51), the entropy value remains the same.
- All rules (strong) of Category-I are characterized by the zero DE, SCI, and SCF values with very high SCO values.

Category-II

- Entropy (BiEntropy values) stabilizes (string either becomes all zero's or a well-observed pattern) from initial high or fluctuating entropy values.
- Rules with a strong association of Category-II characterized by the low to high DE and very high SCO values with zero and one value of SCI and SCF values respectively, whereas, for weak associations, DE and SCO values are usually high with low SCI and SCF values.

Category-III

- Entropy (BiEntropy values) values fluctuate or result in a periodic behavior (BiEntropy values usually remain high) from initial high or fluctuating entropy values.
- All Rules (strong and weak) of Category-III are characterized by the low to high DE and very high SCO values with low to medium SCI and SCF values values (except for Rule 1 and 33, wherein pretty high, e.g., 79, SCI and SCF values alternates)

Category-IV

- Entropy (BiEntropy values) exhibit chaotic beahiour with smaller cell sizes (value of n), which stabilizes as n increases (i.e., $n = 32$).
- All Rules of Category-IV are characterized by the low to high DE and high SCO values with very low SCI and low to high SCF values.

Category-V

- Rules in Category-V exhibit mixed behavior (of previous categories) with different values of cell length (value of n).
- None of these rules showed a zero value (zero indicates a high level of overlapping) of DTW analysis with any other rule.

Table 1: Category-wise variation in the four measurements (i.e., DE, SCO, SCI, and SCF)

Category	DE	SCO	SCI	SCF
I	0	very high	0	0
II	low to high	very high	0	1
III	low to high	very high	low to medium	low to medium
IV	low to high	high	low	low to high
V	Mixed			