Proceedings of the 18th Conference on Computer Science and Intelligence Systems pp. 63–64
DOI: 10.15439/2023F0004
ISSN 2300-5963 ACSIS, Vol. 35

Abstract Approach to Entropy and Co-Entropy in Measurable and Probability Spaces Invited Lecture—Extended Abstract

Gianpiero Cattaneo*
* Retired from Department of Informatics, Systems and Communications
University of Milano–Bicocca, Italy

THE talk is divided in following parts.

(Pa1) First of all, the talk is based on the *Universe* $\mathcal{U} := (\mathbb{N}, \mathcal{P}(\mathbb{N}))$, consisting of the set of natural integers $\mathbb{N}$ and its power set $\mathcal{P}(\mathbb{N})$. Elements $a \in \mathbb{N}$ are interpreted as *micro-states* of the system and subsets $A \subseteq \mathcal{P}(\mathbb{N})$ of the universe are interpreted as *events* which can be tested on it. The first part concerns the abstract approach to *measure distributions of total measure* $M > 0$, as non negative sequences of integer numbers $\widehat{m} := (m_1, m_2, \ldots, m_M)$, i.e., with any $0 \leq m_j \in \mathbb{N}$, and satisfying the condition $\sum_{j=1}^{\infty} m_j = M$. These measure distributions can also be interpreted as *physical macro-states* of the system and their collection will be denoted as $\mathcal{MD}_M$.

From any M-measure distribution $\vec{m}$ it is induced, in a standard way, the *probability distribution* $\vec{p} := (p_1 = m_1/M, p_2 = m_2/M, \ldots, p_M = p_M/M)$, where any $p_j := m_j/M \geq 0$ and $\sum_{j=1}^{M} p_j = 1$ (from which it follows that $p_j \in [0,1]$, according to the usual definition of probability). Based on these basic notions, the following important definitions are then introduced:

a) The *entropy* associate to a M-measure distribution $\vec{m}$ is defined as $H(\vec{m}) := \log M - \frac{1}{M}\sum_j m_j \log m_j$ (from which it follows the standard Shannon entropy $H(\vec{p}) = -\sum_j p_j \log p_j$).
b) The *co-entropy* (i.e., *complementary* entropy) of $\vec{m}$ is defined as $K(\vec{m}) := \frac{1}{M}\sum_j m_j \log m_j$ (from which it follows $K(\vec{p}) = \sum_j p_j \log(M \cdot p_j)$).

Now, once introduced the *total entropy* as $E(\vec{m}) := H(\vec{m}) + K(\vec{m})$, it is easy to prove the *total entropy conservation principle*: $\forall \vec{m}$, $E(\vec{m}) = \log M$ (resp., $\forall \vec{p}$, $E(\vec{p}) := H(\vec{p}) + K(\vec{p}) = \log M$), quantity *always* equal to the constant $\log M$, i.e., invariant with respect to the considered measure distribution $\vec{m}$ (resp., to the considered probability distribution $\vec{p}$).

(Pa2) In the context of the *concrete Universe*, the second part is dedicated to the partition $\pi(\mathbb{N}) = \{A_1, A_2, \ldots, A_K\}$ ($K \leq M$) of the universe $\mathbb{N}$, formed by the so-called *blocks* A_j, and generating the equivalence relation $\equiv$ on $\mathbb{N}$:

let $a_r, a_s \in \mathbb{N}$, then $a_r \equiv a_s$ iff $\exists A \in \pi(\mathbb{N})$: $a_r \in A$ and $a_s \in A$.

The partition $\pi(\mathbb{N})$ determines the measure distribution:

$$\vec{m}(\pi(\mathbb{N})) = \big(m(A_1) = |A_1|, \\ m(A_2) = |A_2|, \ldots, m(A_K) = |A_K|\big),$$

of total measure $m(\pi(X)) := \sum_{j=1}^{K} m(A_j) = \sum_{j=1}^{K} |A_j| = M$; with corresponding entropy and co-entropy.

But in some sense, the present part (Pa2) constitutes a point of connection between the abstract content of point (Pa1) and the concrete part which constitutes the fundamental content of the next point (Pa3).

(Pa3) The collection $\mathcal{MD}_M$ of all measure distributions was the subject of a long research conducted together with French collaborators many years ago entitled *Ice Pile Models*. Brylawski in a paper published od Discrete Mathematics (1973) investigated from a pure algebraic point of view the collection $\mathcal{IP}_M$ of M-uples of non-negative integers $\alpha := (a_1, a_2, \ldots, a_M)$, called *integer partitions*, satisfying the further condition of being *non-increasing*: $\forall i$, $a_{i-1} \geq a_i$. Trivially, $\mathcal{IP}_M \subseteq \mathcal{MD}_M$. This collection has a lattice structure with respect to a lattice order $\leq$, called *dominance*, with minimum element $\mathbf{0} := (M, 0, \ldots, 0)$ and maximum element $\mathbf{1} := (1, 1, \ldots, 1)$, i.e., $\forall \alpha \in \mathcal{IP}_M$, $\mathbf{0} \leq \alpha \leq \mathbf{1}$. On the lattice $\mathcal{IP}_M$ Brylawski introduces the definition of *elementary transition* as the binary relationship: $\alpha \to \beta$ iff $\alpha \lneqq \beta$ and $\exists \gamma$: $\alpha \leq \gamma \leq \beta$, then either $\gamma = \alpha$ or $\gamma = \beta$. On the standard configuration space $\mathcal{IP}_M$ of Brylawski, we interpret the component of place i of the configuration α as a column of a_i grains that can move from left to right. The most interesting result is the following

Theorem – *The configuration transition $\alpha \to \alpha'$ in the space $\mathcal{IP}_M$ occurs iff one of the only two cases occurs: (1) The* vertical *evolution rule (VR) which solves jumps in the configuration of two units. Formally,*

$$\textit{let } a_{j-1} \geq a_j + 2, \quad \textit{then} \\ (a_{j-1}, a_j, a_{j+1}) \xrightarrow{(VR)} (a_{j-1} - 1, a_j + 1, a_{j+1})$$

(2) the horizontal evolution rule (HR) in which a grain can slip from column $j-1$ to column j according to the following local behaviour

$$\text{let } a_{j-1}-1=a_j, \quad \text{then} \quad (a_{j-1},a_j,a_{j+1}) \xrightarrow{(HR)} (a_{j-1}-1,a_j+1,a_{j+1})$$

But, the Ice Pile model has been defined by Goles–Kiwi (1993) just as the structure based on the Brylawski lattice equipped with the two local evolution rules (VR) and (HR). Furthermore, from another point of view, since the two local rules involve neighborhoods of the center a_i of radius 1, they define a one-dimensional *Elementary Cellular Automata*. In any case, whatever the point of view adopted, the application of the two rules (HR) and (VR) determine a discrete time dynamical system since it is possible to consider all the dynamic evolutions $\gamma : \mathbb{N} \mapsto \mathcal{IP}_M$ of initial state $\mathbf{0}$ of the kind $\mathbf{0} \to \alpha(1) \to \alpha(2) \to \ldots \to \alpha(t) \to \ldots$, which by a Theorem converges after a finite number t_f of time steps to the unique final equilibrium configuration $\ldots \to \alpha(t_f) = \mathbf{1}$.

The important point is that if to any configuration $\alpha(t)$ of a trajectory γ one calculates the corresponding entropy $H(\alpha(t))$ one obtains the strictly increasing chain of positive numbers $0 = H(\mathbf{0}) < H(\alpha(1)) < H(\alpha(2)) < \ldots < H(\alpha(t)) < \ldots < H(\mathbf{1})$ and, also important result, if any transition $\alpha(t) \to \alpha(t+1)$ is the result of a parallel application of the (VR) rule to $\alpha(t)$, then the dynamical evolution is unique. It is an open problem to prove some similar result in the case of the Ice Pile parallel dynamical evolution.

(Pa4) The forth part is dedicated to the exposition of those theories which are a concrete application of the abstract treatment made in point (Pa1) of entropy in measurable spaces:

(1) the *Pawlak rough set theory* as mathematical approach to imperfect knowledge, with its application to information systems, based on a non-empty set X of objects forming the Universe of the discourse, a non-empty set Att of attributes, and a function $F : X \times Att \mapsto val$ assigning to any pair $(x,a) \in X \times Att$ consisting of an object x and an attribute a a value $F(x,a) \in val$. Fixed a collection $\mathcal{A} := \{a_1, a_2, \ldots, a_N\}$ from the set of all attributes Att, two objects x_1 and x_2 are considered *indistinguishable* with respect to $\mathcal{A}$ iff $F(x_1,a) = F(x_2,a)$ for every attribute $a \in \mathcal{A}$. Trivially, this binary relation of indistinguishability is an equivalence relation on the set of all objects X, inducing a partition $\pi_{\mathcal{A}}(X)$ of the universe X, and so all the results of part (Pa2) can be applied to the present case of rough set theory;

(2) the *Zadeh fuzzy set theory*, as distributive lattice equipped with a (unique) Kleene negation connective. The De Luca–Termini approach to fuzzy set as distributive lattice equipped with a (unique) Brouwer, or intuitionistic, negation connective has been successively introduced. Based on these two (non intersecting) approaches, it is then defined the Brouwer–Zadeh (BZ) distributive lattice structure equipped with both the Zadeh and the Brouwer negation connectives, where rough sets constitute its crisp part (without however capturing all the applicative richness exposed in point (1)).

(3) the *conservative self–reversible logical gates*. Conservative logic is a model of computation whose principal aim is to compute with zero dissipation of internal Shannon entropy. This goal is reached by basing the model upon reversible and conservative primitives, for example Fredkin and Toffoli gates, which reflect physical principles such as the *reversibility* of microscopic dynamical laws and the *conservation* of certain physical quantities, such as the entropy of the physical system used to perform the computations.

Finally, a standard procedure for embedding non-reversible logical gates into reversible–conservative logical ones is exposed.