A Note on Complexity Measures
for Probabilistic P Systems

Andrés CORDÓN-FRANCO, Fernando SANCHO-CAPARRINI

Research Group on Natural Computing
Department of Computer Science and Artificial Intelligence
University of Sevilla
Avda. Reina Mercedes s/n, 41012 Sevilla, Spain
E-mail: `fsancho@us.es, acordon@us.es`

Abstract. In this paper we present a first approach to the definition of different entropy measures for probabilistic P systems in order to obtain some quantitative parameters showing how complex the evolution of a P system is. To achieve this, we define two possible measures, the first one to reflect the entropy of the P system considered as the state space of possible computations, and the second one to reflect the change of the P system as it evolves.

1   Introduction

In [3] Gh. Păun introduced a new computing device, called P system, within the framework of *Natural Computing* based upon the observation that the processes which take place in the complex structure of a living cell can be considered as *computations*. Since their definition, these complex systems were investigated from many points of view. For instance, they were shown to be able to solve *hard* problems in lower time than classical devices. Starting from an initial state, they *evolve* along the execution by means of rules that can modify the objects inside the system and their own structure.

If we look at P systems as probabilistic devices where several different computations can be reached from a given initial state, it is intuitively obvious that some uncertainty in the result provided by the P system arises. In order to define some quantitative property allowing us to measure this uncertainty, we begin with some general observations about the notion of a system.

A common-sense definition for the notion of a *system* is *a group of units so combined as to form a whole and to operate in unison.* There are several other definitions in the literature. For example, a technical definition was presented by Hall and Fagan [1]: *A system is a set of objects together with relationships between the objects.* Formalists present a very simple and general definition: given a family of base sets $X_1, \ldots, X_n$, then a system S is any relation (subset) in $X = \Pi_{i=1}^n X_i$. Constructivists disagree with this formalist point of view, emphasizing that the natural world of evolving systems can never be captured by formal systems with an *a priori* fixed and finite amount of base sets; hence, they propose *open systems* which define their elements and base sets during the processes of their evolution.

Nevertheless, there are some common characteristics in all above approaches:

443

- A variety of distinct entities.

- These entities are involved in some kind of relations.

- These relations are sufficient to generate a new entity in a higher complexity hierarchy.

In the last 20 years, the study of *complex systems* has emerged as a recognized field in its own right, although a good definition of what a complex system is has eluded rigorous formulation. Attempts to formalize the concept of *complexity* go back even further, to Shannon's *Information Theory* [5], where the concept of *entropy* and *information* of a system (as a measure of uncertainty and its variety) was defined.

2   Entropy and Information

The *state space*, S, of a system is the set of all possible states that the system can be in. An essential component in the study of a system is a quantitative measure for the size of its state space. This measure is usually called *variety*. This measure represents the freedom of the system to reach a particular state, and thus the uncertainty we have about which state the system is in. The variety, V, is defined as the number of elements in the state space or, more commonly, as $V = \log_2(|S|)$, if we encode the states of the system by bits. A variety of one bit, $V = 1$, means that the system has two possible states, that is, one choice. In the case of n binary variables, $V = \log_2(2^n) = n$ is therefore equal to the number of independent choices.

A system's variety, V, measures the number of possible states the system can exhibit, and corresponds to the number of independent binary variables. But in general, the variables used to describe a system are neither binary nor independent. If the actual variety of states that the system can exhibit is smaller than the variety of states we can potentially conceive, then we say that the system is *constrained*, what means that the system cannot make a full use of the *available freedom*, because some *internal or external laws* don't allow certain combinations of values for the variables. Constraint reduces our uncertainty about the system's state.

Hence, the *constraint*, C, of the system can be defined as the difference between maximal and actual variety:

$$C = V_{max} - V$$

The *variety* and *constraint* can be generalized to a probabilistic framework, where they are replaced respectively by *entropy* and *information*.

Let us suppose that we do not know the precise state, s, of a system, but only the probability distribution, $P(s)$, of the system to be in state s. The generalization of the variety of the system can then be expressed as entropy H:

$$H(P) = -\sum_{s \in S} P(s) \cdot \log(P(s))$$

H reaches its maximum value if all states are equiprobable, that is, if we have no information to assume that one state is more probable than another state (and, in this case, entropy reduces to variety). Like variety, H expresses our uncertainty or ignorance about the system's state. The following result, where we obtain maximal certainty (or complete information) about the state of the system, is straightforward.

Lemma 2.1 *The entropy vanishes, i.e., $H = 0$, if and only if there exists a state, $s \in S$, such that $P(s) = 1$.*

As we have seen, the constraint reduces the uncertainty, that is, the difference between maximal and actual uncertainty. This difference can also be interpreted in a different way, as *information*, and historically H was introduced by Shannon [5] as a measure of the capacity for information transmission of a communication channel. If we achieve information about the state of the system, then this information will reduce our uncertainty about the system's state, by excluding (or reducing the probability of) a number of states. The information I we receive from this achievement is equal to the *amount* of uncertainty that is reduced, that is, the difference between the previous knowledge about the system and the later one,

$$I = H_{before} - H_{after}$$

Although Shannon disagreed with the use of the term *information* to describe this measure, his theory came to be known as *Information Theory*. From then, entropy has been used as a measure for a number of higher-order relational concepts, including complexity and organization, and it has been applied in several knowledge fields (as biology, ecology, psychology, sociology, and economics) where the use of complex systems is unavoidable.

3   Measuring the Entropy of Probabilistic P Systems

At this point there is no doubt that a probabilistic P system, PPS for short (no mind the way to define the probability *controlling* the evolution of the system) can be a complex system with the possibility to evolve along the time. Hence, all above parameters can be applied to measure the complexity of this kind of systems.

Irrespective which is the chosen possibility to introduce probabilities in membrane system (see [2] for a detailed study), if we look at the computations the system generates, the result is a tree with a probability measure over its nodes.

Let us remember that over a rooted tree we can define a direct relation between adjacent nodes. With respect to this relation, we denote, for every node x of the tree, by $Ch(x)$ the set of its children (that could be empty if x is a leaf; we denote by L the set of leaves of the tree).

Definition 3.1 *Given a rooted tree G, with root r, we say that a function $P : V(G) \to [0,1]$ is a* probability function *over G if the following conditions are verified:*

- $P(r) = 1$,

- $\forall x \in V(G) \ (P(x) > 0)$,

- $\forall x \in V(G) \ (x \notin L \to \sum_{y \in Ch(x)} P(y) = 1)$.

In this context, a PPS generates a rooted tree with a probability associated with the nodes labelled by configurations reachable in the computation, where the root of the tree is the initial configuration of the system, and the probability of every node is the probability to reach it from the previous configuration; it will be denoted by $\mathbf{T}(\Pi)$ (or briefly by $\mathbf{T}$ if there is no confusion). The set of maximal branches of $\mathbf{T}$ (the computations of the P system) will be denoted by $Comp(\Pi)$. If $\mathcal{C} \in Comp(\Pi)$, then we denote by $\mathcal{C}_i$ the i-th

configuration of the computation $\mathcal{C}$, therefore $\mathcal{C}_0$ is the initial configuration of the P system (see [4] for a formal definition of these concepts).

Because the P system is a dynamic system where the probability of different evolutions depends on the actual state of the device, in order to capture this evolution, and reflect the instant entropy of it, we propose two different entropies, one of them a *global* one, and one of them a *dynamic* one.

In the case of the global entropy, we consider the set of possible computations of the PPS as the state space, that is, we study the different *final* states the P system can reach in its whole execution.

To define the entropy for the global case, we need to *define* a probability measure over $Comp(\Pi)$. We achieve this by using the probability defined over the PPS:

Definition 3.2 *Given a PPS, (Π, P), we define a new probability measure over $Comp(\Pi)$, that will be denoted with the same symbol, as:*

$$P(\mathcal{C}) = \lim_{n \to |\mathcal{C}|} \prod_{i \leq n} P(\mathcal{C}_i)$$

Note 3.1 *It is easy to check that above definition is, in fact, a probability measure over $Comp(\Pi)$.*

According to the previous definition of entropy and the probability defined in $Comp(\Pi)$, the global entropy of the PPS is the following one.

Definition 3.3 *The* global entropy *of a PPS, (Π, P), is defined as:*

$$H_g(\Pi) = - \sum_{\mathcal{C} \in \; Comp(\Pi)} P(\mathcal{C}) \cdot \log(P(\mathcal{C}))$$

Note 3.2 *Since $\lim_{x \to 0} x \cdot \log(x) = 0$, we consider $0 \cdot \log(0) = 0$.*

Note 3.3 *This quantity measures the uncertainty of the P system Π to evolve along the possible computations. Note that, if the P system has only one possible computation, then it will have the probability equal to 1, hence, from lemma 2.1, the entropy of the system will be 0; that is, there is no uncertainty in the evolution of the system.*

But, is there any way to define a measure reflecting the instant entropy of the system? Next, we will try to define a more complex measure for the entropy of the PPS, where in some way the instant of the execution of it will be captured. But, previously we need some definitions over trees.

In a rooted tree, G, we can consider a *depth* function defined recursively as follows:

$$dep(x) = \begin{cases} 0 & \text{if } x \text{ is the root of } G, \\ dep(f(x)) + 1 & \text{otherwise.} \end{cases}$$

where for every node x not being the root, let us denote by $f(x)$ the father of x in G. The *total depth of G* is defined as

$$dep(G) = \max_{x \in G} \; dep(x)$$

For every $n \leq dep(G)$, we define the *n-level of G* as $G_{|n} = \{x \in G : dep(x) = n\}$, the *leaves of G* as $L = \{x \in G : Ch(x) = \emptyset\}$, $L_n = L \cap G_{|n}$, and

$$G_n = G_{|n} \cup \bigcup_{k<n} L_k$$

If $x \in G$, then we define the *path to x* as the set of nodes of G connecting the root of the tree with the node x (there is only one set verifying this condition, because G is a tree), and it will be denoted by γ_x.

If (G, P) is a tree with a probability function defined over its nodes, by using the paths defined above, we can define a new probability function (derived from P) over the levels of G.

Lemma 3.1 *Let (G, P) a tree with a probability function defined over its nodes. Then, for every $n \in \mathbf{N}$ the following function is a probability function over G_n:*

$$\forall x \in G_n, \ P_n(x) = \prod_{y \in \gamma_x} P(y)$$

If (Π, P) is a PPS, then $\mathbf{T}$ is a rooted tree, so we can apply all above definitions, obtaining the *n-level entropy* of the system.

Definition 3.4 *For every $n \in \mathbf{N}$, the n-level entropy of a PPS, (Π, P), is defined as:*

$$H_n(\Pi) = - \sum_{C \in \mathbf{T}_n} P_n(C) \cdot \log(P_n(C))$$

Note 3.4 *The sequence $\{H_n(\Pi)\}_{n \in \mathbf{N}}$ gives a dynamic measure of the evolution of the PPS Π that approximates the global entropy above defined.*

Lemma 3.2 *For every PPS, Π, it is verified that $\{H_n\}$ is a monotonic growing sequence and:*

$$\lim_{n \to \infty} H_n(\Pi) = H_g(\Pi)$$

Now we can define the amount of information of the P system Π along its execution.

Definition 3.5 *The sequence of* information *of the PPS Π is defined as*

$$I_n(\Pi) = H_n(\Pi) - H_{n+1}(\Pi)$$

4 Conclusions

References

[1] A.D. Hall, R.E. Fagan. Definition of a System. *The General Theory of Systems Applied to Management and Organization.* D.M. Jamieson, G. Chen, L.L. Schkade and C.H. Smith (Eds.). Seaside, Ca.: Intersystems Publications, 1980, 217–243.

[2] A. Obtulowicz, Gh. Păun. (In search of) Probabilistic P systems, *BioSystems*, **70**, 2 (2003), 107–121.

[3] Gh. Păun. Computing with membranes, *Journal of Computer and System Sciences*, **61** 1, (2000), 108–143.

[4] M.J. Pérez-Jiménez, F. Sancho-Caparrini. A formalization of basic P systems, *Fundamenta Informaticae*, **49**, 1–3 (2002), 261–272.

[5] C.E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, **27**, 379–423 and 623-656, July and October, 1948.