Linguistic entropy of an ecological model

Y. Villacampa¹, J. L. Usó², M. A. Castro¹ & P. Sastre³
¹Departamento de Análisis Matemático y Matemática Aplicada, E.P.S.A., Universidad de Alicante, 03080 Alicante, Spain
²Departament de Matemàtiques, Campus de Peryeta Roja, Universitat Juame I, Castellón, Spain
³Departamento de Matemáticas, Universidad Nacional del Centro de la Provincia de Buenos Aires, 7300-Azul, Argentina

Abstract

The mathematical models and the ecological models as a particular case, are a cognitive structure make up of subsystems (words, sentences). We can consider relations between classes of components (syntax) leading to the style and organisation (grammar).

The linguistic units of ecological models (variables, transformed functions, etc.) have a probability to appear in a model and moreover they have certain capacity to combine themselves.

In accordance with the methodology used by the authors in the study of modelling in the construction of mathematical equations, a set of transformed functions should be chosen. Starting from these functions a language in which the equations are written is generated.

It is important to make a study of the entropy of each equation and of its resulting information, because this will allow us to determine the most suitable set of functions.

In this paper the authors study important aspects of the model language. We will define the entropy and quantity of information of the mathematical equations and the entropy and quantity of information of differential system equations that describe a system.

Several linguistic entropies will be used to compare different sub-models of a model with different models, since it will be possible to determine information measurements.
1 Linguistic entropy of Ecological Models

Let be an ecological system and a model of it. When we consider a model for an ecological system we use a math language where ecological processes of this system are modelled, $F_D$. This set of ecological processes are expressed by a set of variables, $\{y_i\}_i$, which are written in different forms through expressions in a math language previously defined, using a vocabulary. If we take several modelling of these processes, we will have several texts written in the same language. Then it is interesting to study their linguistic entropy and then the quantity of information of these. On the one hand, it will allow us choose the best model and on the other hand it will lead us to study the quantity of information of the model.

1.1 Linguistic Entropy of an Ecological Process

Let $F_D = \{M_\alpha, \alpha \in I\}$ be the set of all words that we can obtain from a vocabulary, $V_D$, where the processes defined in an ecological model are modelled. We stand for $y_i$ an ecological process.

It is necessary to know the probability of the words used in the modelling of a process to define its entropy. If we suppose that the process $y_i = \sum_{j=1}^{n} A_i^j M_j$, then the definition of the entropy of $y_i$ is determinate if we know the probability of each $M_j$, $p(M_j) = p_j$. There are several methods to determine this probability. Because of the ecological system nature and their modelling, the modeller will have experimental data, initially. These data can be real or else can be simulated data. So, a process $y_i$ can be worded by different expressions that we will write by $\{\Psi_k\}_{k}$. For this reason the authors propose a way to compute the probability of each word that appears in the equation $\Psi_j$ chosen between all equations $\Psi_k$ which the modeller knows to model an ecological process mathematically. For a given text, which is formed by sentences whose words are element of $F_D$, it is possible to apply the Zipf’s law to determinate the probability of each word.

1.1.1 Calculus of the probability

We consider a text formed by a set of sentences $\{\Psi_k\}_{k}$ written using words of $F_D$.

$$\Psi_r = \sum_{j=1}^{m_r} A_{r,j} M_{r,j}, \quad M_{r,j} \in F_D, \quad \text{for all } j = 1, 2, \ldots, m_r$$  \hspace{1cm} (1)

We will apply the Zipf’s law to determinate the probability of each word. If we arrange the words by decreasing way according to their frequency of appear, we will express by $n$ the rank of the word that occupy the place $n$ and $p_n$ its probability. Then according to the Zipf’s law, $p_n = \frac{k}{n}$ where
\( k \) is a constant. Moreover we can determinate the numerical value of \( k \) for each text. Let \( l \) be the number of words that appear in the text and \( p_i, i = 1, 2, \ldots, l \) their probabilities. It is true that

\[
1 = \sum_{i=1}^{l} p_i = \sum_{i=1}^{l} \frac{k}{l} = k \left( \frac{1}{2} + \cdots + \frac{1}{l} \right)
\] (2)

then

\[
k = \frac{1}{1 + \frac{1}{2} + \cdots + \frac{1}{l}}.
\] (3)

For \( l = 1 \) we obtain that \( p_1 = k \), that is, \( k \) is the probability of the most frequent word.

**Definition 1:** Let \( \{y\} \) be an ecological process and let \( T_y \) be a text written in a language \( F_D \), formed by several sentences that modelling it and that we express by \( T_y \cong \{\psi_k\}_{k=1}^{m} \).

It is defined the linguistic entropy of the first order of the ecological process \( y \) in the text \( T_y \) as

\[
H_{T_y}^1 = - \sum_{i=1}^{l} p_i \log_2 p_i
\] (4)

where \( p_i \) is the probability of a word, computes according to the Zipf’s law.

### 1.2 Linguistic entropy of two Ecological Processes

We consider:

1. Two ecological processes defined by the variables \( y_i \) and \( y_j \).

2. Two texts, \( T_{y_1} \cong \{\psi_k\}_{k=1}^{m_1} \) and \( T_{y_2} \cong \{\phi_k\}_{k=1}^{m_2} \) which are formed by sentences that model the process \( y_1 \) and \( y_2 \).

Let \( \{M_1, M_2, \ldots, M_{l_1}\} \) and \( \{N_1, N_2, \ldots, N_{l_2}\} \) be the sets of all words that appear in the texts \( T_{y_1} \) and \( T_{y_2} \) respectively.

We assume that the words which appear in each text are “independents” to the effect that if

1. \( p_{ij} \) is the probability of the word \( M_i \) appears in the text \( T_{y_1} \), and the words \( N_j \) appear in the text \( T_{y_2} \).

2. \( p_i \) is the probability of the word \( M_i \) appears in the text \( T_{y_1} \).

3. \( p_j \) is the probability of the word \( N_j \) appears in the text \( T_{y_2} \).
then
\[ p_{ij} = p_i \cdot p_j. \] (5)

**Definition 2:** It is defined the entropy of the first order of the processes \( y_1 \) and \( y_2 \) in the texts \( T_{y_1} \) and \( T_{y_2} \) as

\[ H^1_{T_{y_1}, T_{y_2}} = - \sum_{i=1}^{l_1} \sum_{j=1}^{l_2} p_{ij} \log_2 p_{ij}. \] (6)

**Theorem 1:** The entropy of the first order of the ecological processes \( y_1 \) and \( y_2 \) in the texts \( T_{y_1} \) and \( T_{y_2} \) is equal to the addition of the entropy of the first order of each process in each text, that is

\[ H^1_{T_{y_1}, T_{y_2}} = H^1_{T_{y_1}} + H^1_{T_{y_2}} \] (7)

**Proof.** If we use the definition and the fact that \( p_{ij} = p_i \cdot p_j \), (that is, the words are independent), then we obtain that

\[
\begin{align*}
H^1_{T_{y_1}, T_{y_2}} &= - \sum_{i=1}^{l_1} \sum_{j=1}^{l_2} p_{ij} \log_2 p_{ij} = - \sum_{i=1}^{l_1} \sum_{j=1}^{l_2} p_i \cdot p_j \log_2 p_i \cdot p_j = \\
&= - \sum_{i=1}^{l_1} \sum_{j=1}^{l_2} p_i \cdot p_j (\log_2 p_i + \log_2 p_j) = \\
&= - \sum_{i=1}^{l_1} \sum_{j=1}^{l_2} p_i \cdot p_j \log_2 p_i - \sum_{i=1}^{l_1} \sum_{j=1}^{l_2} p_i \cdot p_j \log_2 p_j = \\
&= - \sum_{i=1}^{l_1} p_i \log_2 p_i \sum_{j=1}^{l_2} p_j - \sum_{i=1}^{l_1} p_i \log_2 p_j \sum_{j=1}^{l_2} p_i = \\
&= H^1_{T_{y_1}} \sum_{j=1}^{l_2} p_j + H^1_{T_{y_2}} \sum_{j=1}^{l_2} p_i = H^1_{T_{y_1}} + H^1_{T_{y_2}}
\end{align*}
\]

because \( \sum_{j=1}^{l_2} p_j = \sum_{i=1}^{l_1} p_i = 1. \)

### 1.3 Linguistic entropy of n Ecological Processes

In a similar way that we defined the entropy of the first order of two processes in two texts, we can define the entropy of the first order of \( n \) processes \( \{y_k\}_{k=1}^n \) in \( n \) texts \( T_{y_k} \). In this case, it is easy prove too that

\[ H^1_{T_{y_1}, T_{y_2}, \ldots, T_{y_n}} = \sum_{i=1}^{n} H^1_{T_{y_i}}. \] (9)
1.4 Linguistic entropy of an Ecological Models

Let be an ecological system and a model of this. We suppose that the ecological processes considered in the model are expressed mathematically in a language $F_D$. If we change the experimental or simulated data, we obtain new sentences that model the ecological processes. That is, we obtain texts written in the language $F_D$, $\{T \gamma_i\}_{i=1}^n$.

We define the entropy of the ecological model according to the changes done as the entropy of the processes in the texts that we obtain. If $\gamma$ express the set of changes we define

$$H_{\gamma}^{1} = H_{T_{\gamma_1},T_{\gamma_2},...,T_{\gamma_n}} = \sum_{i=1}^{n} H_{T_{\gamma_i}}^{1}. \quad (10)$$

2 Conclusion.

In this paper, we have defined the linguistic entropy of an ecological model as for a language $F_D$, where we write the math equations, which model the process. These measurement allow us study the order and the quantity of information of the ecological processes relating to their modelling, which an important tool to the modeller. It will allow study what systems analyse the reality behaviour better. These measurements are the starting point to compare different models of an ecological system, such as different languages and it will allow us obtain information rates in following investigations.

References