

Modelo de patrones geométricos de señales 1D basado en autoestructuras para la construcción del espacio de características de un clasificador

Geometric pattern model of 1D signals based on self-structures for the construction of the feature space for a classifier

Presentación: 4 y 5 de Octubre de 2022

Doctorando:

Hernán Manuel García Blesa

Centro de Procesamiento de Señales e Imágenes - UTN, Facultad Regional Buenos Aires.
hgarciablesa@frba.utn.edu.ar

Directora:

Andrea Rey

Codirector:

Walter Legnani

Resumen

El modelo que presentamos en este trabajo se basa en la formulación de parámetros novedosos para distinguir distintos tipos de señales discretas. Realizamos el cálculo de los parámetros mediante el método de barrido con una dimensión de embebedo - *embedding*, en inglés - utilizado ampliamente en aplicaciones de la Teoría de la Información. En este trabajo proponemos tres parámetros simples, a saber: amplitud, ángulo de Zenit y factor de forma. Con estos parámetros obtenemos una terna para cada segmento de una señal, y para cada señal de un grupo dado. Cada señal genera una cantidad de ternas que luego se reducen a ternas únicas, quitando la información repetida. Aplicamos la propuesta al análisis de señales sintéticas, aleatorias y caóticas. Hemos comprobado que la presente metodología es eficiente en diferenciar las señales estudiadas.

Palabras clave: Amplitud, Ángulo de Zenit, Factor de Forma, Diferenciación de Señales.

Abstract

The model presented in this paper is based on the formulation of novel parameters to distinguish different types of discrete signals. We compute the parameters using the scanning method with a given embedding dimension, which is widely used in Information Theory applications. In this work, we propose three simple parameters, namely: amplitude, Zenith angle and factor of shape. With these parameters, we obtain a 3-tuple for each segment of a signal, and for each signal in a given group. Each signal generates a number of 3-tuples, which are then reduced to single 3-tuple by removing repeated information. We apply the

proposal to the analysis of synthetic, random and chaotic signals. We have proved that the present methodology is efficient in differentiating the signals studied.

Keywords: Amplitude, Zenith Angle, Factor of Shape, Differentiation of Signals.

Introducción

La diferenciación de señales es una tarea necesaria debido a múltiples factores. Permite la detección de propiedades, comportamientos, conductas y otros aspectos importantes de las fuentes emisoras, como pueden ser la identificación de un fenómeno, la detección de una anomalía o la tendencia a largo plazo (Tangirala, 2018). Dentro de las metodologías de estudio de señales, se encuentran los patrones ordinales que son una herramienta ampliamente utilizada dentro de los últimos veinte años, con una enorme cantidad de aplicaciones exitosas en casos de la más variada naturaleza (Keller *et al.*, 2017). Para más detalles sobre esta técnica nos referimos al trabajo de Amigó (2010).

Para la construcción del modelo propuesto en este trabajo, consideramos conceptos básicos como lo son la amplitud, una medida angular y la forma de la señal en términos de variaciones de picos y valles de la misma. La hipótesis sobre los resultados al aplicar este modelo radica en considerar que estas tres propiedades son suficientes para obtener caracterización completa de la señal a modo de su huella digital.

Definición de características alternativas

Sea $S = [a_1, a_2, \dots, a_n]$ una señal discreta. Dada una dimensión de *embedding* d , se definen los siguientes segmentos de longitud d : $S_i = [a_i, a_{i+1}, \dots, a_{i+d-1}]$, para $i = 1, 2, \dots, n - d$. Con el objetivo de caracterizar S , se definen tres parámetros: 1- La *amplitud* de S , utilizando los valores máximo y mínimo de S . 2- El *ángulo de Zenit* que mide la distribución de los valores dentro de S . 3- El *parámetro de forma* capaz de caracterizar la secuencia de picos y valles de la señal.

Amplitud

Denotando como $\min(S)$ y $\max(S)$ a los valores mínimo y máximo de la señal S , respectivamente; calculamos la amplitud de S : $A(S) = \max(S) - \min(S)$.

Ángulo del zenit

Proponemos un ángulo, que denominamos *ángulo de Zenit* y notamos como ϕ , obtenido a partir de la oscilación del promedio de valores de S entre el valor mínimo y máximo de S . Para definir ϕ utilizamos el valor medio p_S de los valores de S con el fin de observar cómo éste se ubica geoméricamente dentro del espacio acotado entre el $\min(S)$ y el $\max(S)$. Consideremos el siguiente intervalo en \mathbb{R}^2 :

$$I_S = [0, 1] \times [\min(S), \max(S)]. \quad (1)$$

Definimos el ángulo ϕ determinado por los segmentos \overline{PM} y \overline{PN} donde $P = (0.5, p_S)$, $M = (1, \max(S))$ y $N = (1, p_S)$. Luego, con el objetivo de comparar valores de ϕ provenientes de diferentes señales, normalizamos el intervalo I_S : $I = [0, 1] \times [0, 1]$ y el promedio p_S :

$$p = \frac{p_S - \min(S)}{\max(S) - \min(S)}. \quad (2)$$

Finalmente calculamos:

$$\phi(S) = \arctan\left(\frac{1-p}{0.5}\right) = \arctan(2-2p). \quad (3)$$

Dado que el promedio no puede ser menor que el mínimo ni mayor que el máximo, sus valores oscilarán entre 0 y 1 , determinando que el ángulo $\phi \in [0, 1.1071)$.

Factor de forma

Una señal de un fenómeno real puede contener una cantidad de valles y picos en una determinada secuencia que contribuye a su caracterización. Para la formulación del siguiente parámetro, investigamos una posible manera de extraer una medición de la forma de la señal, con la idea de condensarla en un sólo valor. Para lograrlo, proponemos el siguiente cálculo:

1. Consideramos la secuencia $S = \{a_1, a_2, \dots, a_n\}$.
2. Ordenamos estos datos en orden creciente, obteniendo la secuencia $B = \{b_1, b_2, \dots, b_n\} \subseteq S$, donde $b_i \leq b_{i+1}$ para $i = 1, 2, \dots, n - 1$.
3. Eliminamos los elementos repetidos en B obteniendo la secuencia $C = \{c_1, c_2, \dots, c_m\} \subseteq S$, donde $m \leq n$ y $c_i < c_{i+1}$ para $i = 1, 2, \dots, m - 1$.
4. Definimos el vector $O = \{o_1, o_2, \dots, o_n\}$ como $o_i = j$ si $a_i = c_j$, para $i = 1, 2, \dots, n$.

Al generar el vector de números de orden O , es importante notar que si hay valores repetidos en la secuencia bajo análisis, sus respectivos números de orden también se repetirán. Esta particularidad implica que las componentes del vector de orden están acotadas entre 1 y la longitud de la secuencia original. El vector O representa la forma de la secuencia en términos de picos y valles a partir de números enteros positivos.

Con el objetivo de cuantificar y sintetizar este vector, construiremos un número utilizando las componentes de O como exponentes de una base de números primos. Explícitamente, sea $P_n = \{q_1, q_2, \dots, q_n\}$ el conjunto de los primeros n números primos, definimos el valor que representa unívocamente el parámetro de forma como:

$$F(S) = \prod_{i=1}^n q_i^{o_i}. \quad (4)$$

El Teorema Fundamental de la Aritmética garantiza la factorización del parámetro de forma para recuperar la estructura de valles y picos de la secuencia.

Vector de características

A partir de los parámetros ya definidos, dada una señal S de longitud n y tomando un *embedding* d , podemos asociar a cada segmento S_i una terna definida por $(A(S_i), \phi(S_i), F(S_i))$. Con el fin de simplificar notación, indicamos $A_i = A(S_i)$, $\phi_i = \phi(S_i)$ y $F_i = F(S_i)$. Reemplazando en la expresión anterior, definimos el vector $v_i = (A_i, \phi_i, F_i)$. El conjunto de todos estos vectores es $V = \{v_i : i = 1, 2, \dots, n - d\}$. Observar que el cardinal de V es $\#V = \#S - d + 1$.

Con el objetivo de reducir la información redundante –en tanto sea posible– quitamos los elementos repetidos del conjunto V . Consideramos así el conjunto $V' = \{v_1, v_2, \dots, v_k\}$ donde $v_i \in V$ para $i = 1, 2, \dots, k$ y para todo i, j entre 1 y k tales que $i \neq j$, se satisface que $v_i \neq v_j$. Este conjunto, formado por vectores únicos, tiene cardinal $\#V' \leq \#V$.

Es relevante conocer la frecuencia en V de los vectores de V' . Para $i = 1, 2, \dots, k$ definimos la frecuencia:

$$F_i = \#\{v \in V : v = v_i\}. \quad (5)$$

Finalmente, definimos el conjunto

$$W = \{(v_i, F_i) : i = 1, 2, \dots, k\}, \quad (6)$$

que contiene los vectores de V' y sus respectivas frecuencias en V .

Análisis de casos triviales

Para un primer análisis, consideramos los siguientes tipos de señales:

- **Señal constante:** Si S es constante, entonces todos los elementos en V resultan iguales. Luego, el conjunto V' tiene un cardinal $\#V'=1$ y una única frecuencia, que es máxima, dada por $F=\#S-d+1$. De este modo, W está formado por un único elemento que resulta la mínima representación de S .
- **Señal aleatoria:** Si S es una señal aleatoria, significa que el conjunto de vectores únicos V' tiene un cardinal máximo $\#S-d+1$. Dado que los vectores de V' son todos diferentes, la frecuencia de éstos en V será mínima; es decir, $F_i=1$ para todo i . Finalmente, W tiene cardinal igual a $\#S-d+1$, con lo que no se puede reducir la información de S .

Casos de estudio

El modelo se aplicó en una primera prueba a dos grandes grupos de señales sintéticas para estudiar la performance de los tres parámetros descriptos al diferenciar señales. Para ello consideramos un grupo de señales aleatorias uniformes (Shynk, 2012) versus uno de señales caóticas (Devaney, 2020) utilizando la función logística $f(x) \Leftrightarrow \lambda x(1-x)$, con $\lambda \in [3.65, 3.7]$ (para esta implementación nos referimos al trabajo de Lynch (2018)). Ambos grupos contienen 500 señales de 4000 puntos cada una. En una siguiente etapa, se replicó la prueba con 8 grupos de señales definidas mediante funciones elementales, como lo son: seno, coseno, tangente, sigmoidea, cuadrática, cúbicas, raíz y racional.

Resultados

Las Figuras 1 y 2 muestran, respectivamente, los *boxplots* de los tres parámetros junto con el de la frecuencia para el análisis de las señales aleatorias versus caóticas, y para las señales generadas por funciones elementales.

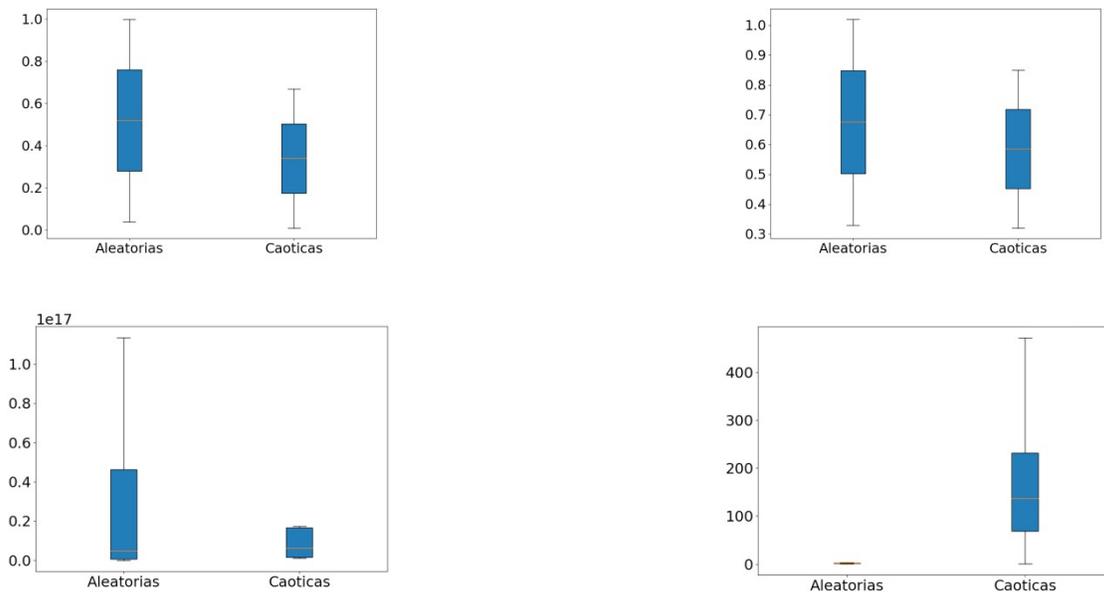


Fig.1: *Boxplots* de los parámetros aplicados a las señales aleatorias y caóticas. De izquierda a derecha, y de arriba hacia abajo: amplitud, ángulo de Zenit, factor de forma y frecuencia.

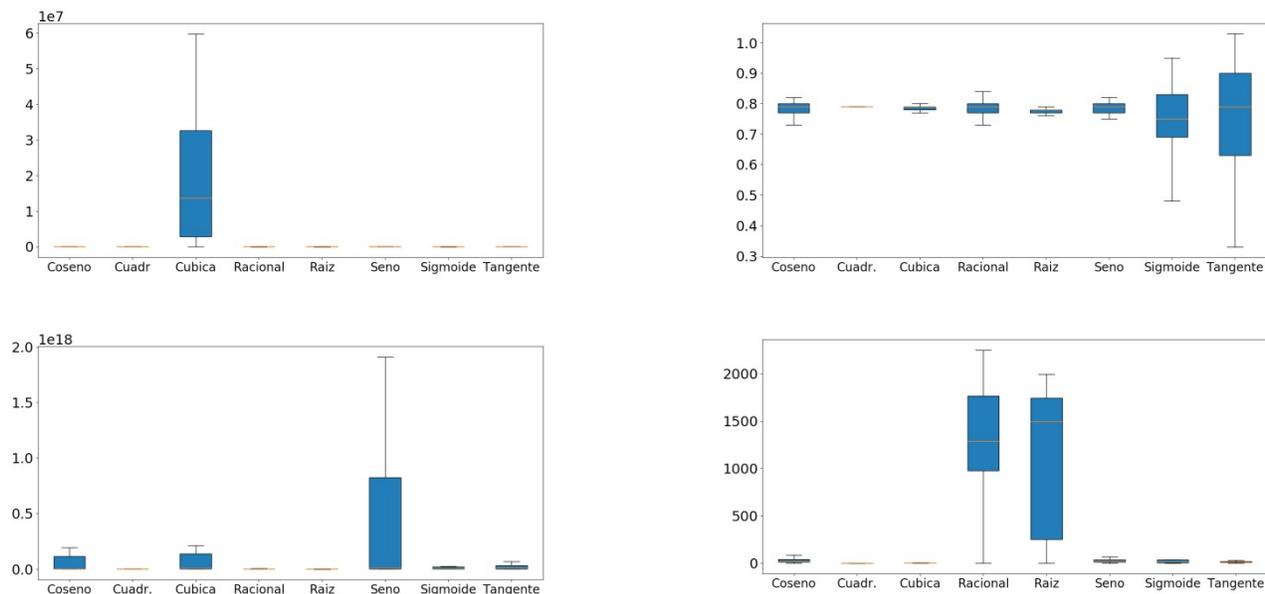


Fig.2: *Boxplots* de los parámetros aplicados a las señales generadas por funciones elementales. De izquierda a derecha, y de arriba hacia abajo: amplitud, ángulo de Zenit, factor de forma y frecuencia.

Conclusiones

Podemos ver que los parámetros definidos poseen características distintivas entre los diferentes grupos de señales analizados. En el caso de las señales aleatorias y caóticas, la frecuencia muestra ser el parámetro que mejor distingue ambos grupos. No obstante, la amplitud también muestra diferencias en los grupos, pero en menor grado. En el caso del estudio de las señales generadas por funciones elementales, el ángulo de Zenit y la frecuencia evidencian fuertes diferencias entre los grupos. El método muestra que, para diversos grupos de señales, ciertos parámetros resultan más sensibles que otros. En particular, podemos decir que el modelo es capaz de diferenciar señales caóticas de señales aleatorias uniformes.

Referencias

- Amigó, J. (2010). *Permutation complexity in dynamical systems: Ordinal patterns, permutation entropy and all that*. Springer Science & Business Media.
- Devaney, R. (2020). *A first course in chaotic dynamical systems: Theory and experiment*. Chapman and Hall/CRC.
- Keller, K., Mangold, T., Stolz, I., Werner, J. (2017). Permutation entropy: New ideas and challenges. *Entropy*, 19(3), 134.
- Lynch, S. (2018). *Dynamical Systems with Applications using Python*. Birkhäuser.
- Shynk, J. J. (2012). *Probability, random variables, and random processes: Theory and signal processing applications*. John Wiley & Sons.
- Tangirala, A. K. (2018). *Principles of system identification: Theory and practice*. Crc Press.