

Clasificación de electrodomésticos a partir del monitoreo de consumos domiciliarios utilizando máquinas de soporte vectorial

Juan L. Barberia, Juan F. Fernández Biancardi, Giancarlo Lottero, Marco G. Baldiviezo, Sebastián L. Esangui Garagoli, Lorenzo Licata Caruso y Walter E. Legnani

Universidad Tecnológica Nacional, Facultad Regional Buenos Aires, Centro de Procesamiento de Señales e Imágenes (CPSI), Medrano 951, Ciudad Autónoma de Buenos Aires, Argentina

glottero@frba.utn.edu.ar

Recibido el 10 de junio de 2023, aprobado el 10 de julio de 2023

Resumen

El monitoreo de carga no intrusivo es un conjunto de técnicas cuyo objetivo es la identificación de los electrodomésticos utilizados en una vivienda partiendo de la señal de consumo energético agregada que producen al utilizarse. La base de datos empleada es la denominada *Reference Energy Disaggregation Data Set* (REDD) la cual es de acceso gratuito y contiene información detallada del consumo eléctrico de artefactos de varios hogares. El presente trabajo tuvo como objetivo utilizar herramientas de la teoría de la información y del análisis de señales para generar un espacio de características como conjunto de entrenamiento para un modelo de *Machine Learning* supervisado (SVM); para formular así una clasificación satisfactoria de electrodomésticos. Los resultados obtenidos muestran una precisión lo suficientemente elevada como para alentar la continuación de esta línea de investigación.

PALABRAS CLAVE: MONITOREO NO INTRUSIVO DE CARGA - MÁQUINA DE SOPORTE VECTORIAL - CONSUMO DOMICILIARIO – ENTROPÍA - PROBABILIDAD DE TRANSICIONES ENTRE PATRONES ORDINALES

Abstract

Non-intrusive load monitoring is a set of techniques aimed at identifying household appliances based on the aggregated energy consumption signal they produce when used. The employed database is called the Reference Energy Disaggregation Data Set (REDD), which is freely accessible and contains detailed information on the electrical consumption of appliances in various households. The objective of this work is to use information theory and signal analysis tools to generate a feature space as a training set for a supervised Machine Learning model (SVM), thus achieving satisfactory appliance classification. The obtained results show sufficiently high accuracy to encourage further continuation of this line of research.

KEYWORDS: NON-INTRUSIVE LOAD MONITORING - SUPPORT VECTOR MACHINES - HOUSEHOLD CONSUMPTION - ENTROPY - ORDINAL PATTERNS TRANSITION PROBABILITY

Introducción

Uno de los mayores retos de la sociedad contemporánea es afrontar y resolver las problemáticas referidas a la eficiencia energética y a la sostenibilidad. La energía mundial proviene, mayormente, de fuentes a base de combustibles fósiles como lo son el petróleo, el carbón y el gas natural. Estas representan más del 82 % de la energía mundial, mientras que las denominadas fuentes de energías renovables representan aproximadamente el 7 %. Las últimas mencionadas engloban las formas de generación del tipo eólica, solar, geotérmica y de biomasa (para más detalle ver Fig. 1). La mayoría de la matriz energética mundial se basa en recursos que no son renovables, por lo que se dispone de una cantidad finita de los mismos y, si se anexa el problema de la combustión masiva de estos recursos, se genera un gran impacto climático-ambiental.

Se suma a lo antes expuesto la ampliación de las llamadas “redes inteligentes” que utilizan flujos bidireccionales de electricidad y de información, teniendo como objetivo crear una red de suministro eléctrico más automatizada. Estas redes permiten desarrollar nuevas aplicaciones vinculadas con el control, gestión y medición de información como lo es el “monitoreo no intrusivo de cargas” o NILM, por sus siglas en inglés “*Non-Intrusive Load Monitoring*”. Esta clase de monitoreo requiere separar las cargas eléctricas examinando sólo las formas de onda del consumo de energía específicas de cada electrodoméstico estudiado dentro de los datos de carga. Los datos se adquieren del tablero eléctrico principal fuera del edificio o residencia, por lo que se considera no intrusivo ya que el método no requiere la instalación de ningún equipo en el interior de la propiedad del cliente.

Un aspecto clave a explorar en las aplicaciones NILM es la adecuada caracterización del consumo de energía eléctrica de los electrodomésticos. Para abordar este tema, en la presente investigación se ha empleado la base de datos del trabajo denominado conjunto de datos de desagregación de energía de referencia o REDD, por sus siglas en inglés, “*Reference Energy Disaggregation Data Set*” (Kolter y Johnson, 2011) y una caracterización del comportamiento de los dispositivos eléctricos mediante el uso de cuantificadores provenientes de la teoría de la información. Por lo que en la investigación convergen tres ejes conceptuales: el monitoreo de consumo domiciliario, herramientas de la teoría de la información y aprendizaje automático.

Base de Datos

Generalidades

La base de datos REDD que fue construida de forma tal que facilite el uso de técnicas de *machine learning* para el estudio de señales de consumo energético de usuarios reales, teniendo en mente la aplicación en investigaciones relacionadas con el monitoreo de carga no intrusivo.

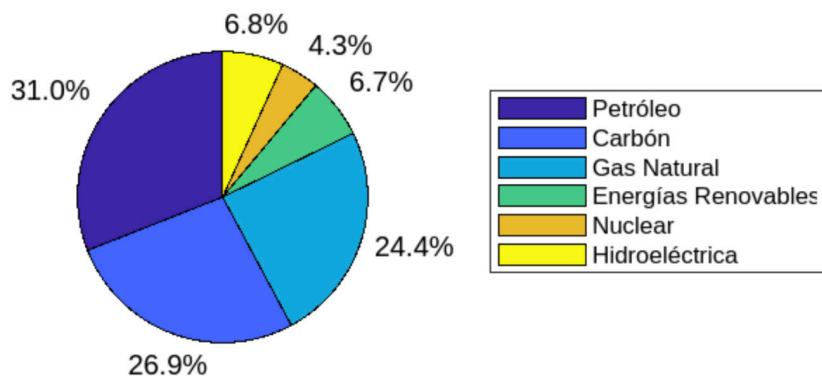


Fig. 1. Matriz Energética Mundial 2021. Adaptado de los datos de (Sánchez Guevara, 2022)

La base de datos empleada contiene información sobre mediciones de energía y se realiza tanto en circuitos individuales como en el conjunto de todos los circuitos. Ofrece un conjunto de datos de potencia muestreados a baja frecuencia y otro de alta frecuencia. Para este trabajo se empleó el de baja frecuencia, que registra cada 3 segundos los circuitos individuales y cada 1 segundo a todo el conjunto.

Las mediciones se tomaron en un total de seis viviendas durante dos meses, registrándose el consumo de un promedio de 20 circuitos por vivienda etiquetados con el nombre del electrodoméstico que lo caracteriza, conformando así las clases a estudiar, aunque no se proporciona información detallada sobre los modelos de electrodomésticos utilizados ni sobre la tecnología de los mismos.

Metodología

Espacio de características

El concepto de entropía emerge acompañando la invención de la máquina de vapor con aplicaciones prácticas durante el siglo XIX, siendo su relación con las señales y la teoría de la información un aporte posterior realizado por Shannon (1948).

En la actualidad, se ha convertido en uno de los principales conceptos para cuantificar la información contenida en una serie de datos, tanto temporales como espaciales. La definición fue extendida luego por Kolmogorov a los sistemas dinámicos y a la complejidad computacional (Kolmogorov, 1958), siendo posteriormente modificada por Sinai (1959) dando lugar al denominado enfoque de Kolmogorov-Sinai. Por otro lado, Olivares y colaboradores (2019), entre otros, proporcionan una serie de interpretaciones del concepto de entropía derivados de una amplia gama de temas científicos y tecnológicos tales como: el desorden en sistemas complejos, la caracterización del volumen del espacio de estado, o la falta de información para predecir el siguiente valor en una serie de datos.

Para formalizar este concepto se puede partir, por ejemplo, de una serie de tiempo:

$$X(t) = \{x_t : t = 1, \dots, T\}, \quad (1)$$

donde

T es el número de valores registrados, la función de densidad de probabilidad (FDP) discreta asociada a X está dada por $P = \{p_i : i = 1, \dots, T\}$,

N representa el número de estados accesibles del sistema en estudio, y que debe satisfacer que $\sum_{i=1}^N p_i = 1$

Continuando con este razonamiento, la entropía de Shannon se encuentra definida por la siguiente expresión:

$$S[P] = - \sum_{i=1}^N p_i \ln(p_i). \quad (2)$$

La expresión (2) puede verse como una medida de la incertidumbre relacionada con el proceso físico descrito por la probabilidad P . De esta forma cuando $S[P]=0$, significa que se dispone del máximo conocimiento del proceso, resultando la estructura subyacente completamente determinista. En presencia de un proceso estocástico, no correlacionado o una distribución uniforme, se presenta la máxima incertidumbre respecto al conocimiento del sistema dinámico, de tal manera que todos los estados del sistema tienen

la misma probabilidad de ocurrencia, la cual puede modelarse por ejemplo mediante una distribución uniforme dada por $P_e = \{p_i = 1/T, i=1, \dots, T\}$ (Sinai, 1959; Brissaud, 2005; Zunino *et al*, 2012).

Un punto que se debe tener muy presente es que, si se dispone de las probabilidades antes mencionadas, no se necesita conocer nada más acerca de la naturaleza del sistema que genera la serie de datos. Es decir, independientemente que se trate de un sistema caótico, estocástico, lineal o no lineal, el tratamiento es el mismo para el cálculo de la entropía. Se suele emplear la entropía normalizada de Shannon, que se define como:

$$H[P] = \frac{S[P]}{S[P_e]}, \quad (3)$$

resultando evidente que $0 \leq H[P] \leq 1$.

Como la entropía de Shannon no es sensible a cambios abruptos de amplitud, se suele emplear en el caso de señales que poseen dichas variaciones la denominada entropía ponderada. En particular en el presente trabajo se empleará la entropía de permutación ponderada H_w (Fadlallah *et al*, 2013). Dada una serie de datos como en (1) se particiona la serie temporal original en vectores de *embedding de dimensión m* de la siguiente manera:

$$X_j = \{x_j, x_{j+\tau}, \dots, x_{j+(m-1)\tau}\} \text{ para } j = 1, 2, \dots, N - (m-1)\tau, \quad (4)$$

siendo τ el denominado *embedding time delay*, que en este trabajo se ha tomado igual a 1 y la dimensión de *embedding m* igual a 4.

A cada uno de estos vectores, se les calcula el valor medio \bar{X}_j dado por (5) y en base a este, se calcula un factor de peso w_j dado por (6).

$$\bar{X}_j = \frac{1}{m} \sum_{k=1}^m x_{j+(k-1)\tau}, \quad (5)$$

$$w_j = \frac{1}{m} \sum_{k=1}^m \left[x_{j+(k-1)\tau} - \bar{X}_j \right]^2. \quad (6)$$

A cada uno de los vectores de *embedding* se les asigna objetos de un alfabeto previamente definido, donde quedan de manifiesto todos los posibles órdenes de las $m!$ permutaciones posibles de las secuencias de dicho vector de *embedding*. Una vez obtenidos, se calcula la frecuencia relativa de estas permutaciones, siendo su cantidad $N = T - (m-1) \cdot \tau$, empleando estas como una estimación de sus correspondientes probabilidades:

$$p_w(\pi_i^{m,\tau}) = \frac{\sum_{j \leq N} 1_{u \text{ tipo}(u) = \pi_i} (X_j^{m,\tau}) w_j}{\sum_{j \leq N} 1_{u \text{ tipo}(u) = \Pi} (X_j^{m,\tau}) w_j}, \quad (7)$$

siendo la función $1_A(u)$ definida como $1_A u = 1$ en el caso de que el vector de *embedding u* pertenezca al objeto del alfabeto A, y $1_A(u) = 0$ si $u \notin A$.

La entropía de permutación ponderada se define entonces de la siguiente manera:

$$H_w^{m,\tau} = - \sum_{i=1}^{m!} p_w(\pi_i^{m,\tau}) \cdot \log_2 p_w(\pi_i^{m,\tau}), \text{ con } p_w(\pi_i^{m,\tau}) \neq 0. \quad (8)$$

Para trabajar con diferentes entropías, es conveniente normalizar la expresión (8) según:

$$H_{wn}^{m,\tau} = \frac{H_w^{m,\tau}}{m!} \tag{9}$$

Para el caso particular de tomar todos los pesos $w_j=1$, la entropía calculada corresponde a la Entropía de Permutación de Bandt y Pompe EP (Bandt y Pompe, 2002), que posee múltiples aplicaciones en bioingeniería (Amigó y Keller, 2013).

Para introducir una definición de complejidad en el campo de la Física, es crucial considerarla como un indicador tanto de la aleatoriedad como de la presencia de estructuras correlacionadas de la serie de tiempo generada por un sistema dinámico dado (Rosso *et al.*, 2010). López-Ruiz y colaboradores (1995) sugieren la inclusión de un tipo de distancia a una FDP conocida o de referencia, para el cálculo de la complejidad. Por lo tanto, el desequilibrio con respecto a dicha distribución de referencia, se define por lo propuesto por (Martín *et al.*, 2006):

$$Q[P] = Q_0 \cdot D[P, P_e] \tag{10}$$

donde Q_0 es una constante de normalización calculada como:

$$Q_0 = -2 \left[\frac{N+1}{N} \ln(N+1) - 2 \ln(2N) + \ln(N) \right]^{-1} \tag{11}$$

y D es una distancia estocástica, que en este trabajo es considerada como la divergencia de Jensen-Shannon dada por:

$$D_{SJ}[P, P_e] = S \left[\frac{P+P_e}{2} \right] - \frac{S[P]}{2} - \frac{S[P_e]}{2} \tag{12}$$

Se puede observar que el desequilibrio se hace nulo cuando la distribución de la serie temporal se asemeja a una nube de puntos uniformemente distribuidos, y que se maximiza cuando la serie temporal es periódica. Así, la complejidad puede medirse como una combinación entre la información inherente al sistema y una estimación del desequilibrio respecto a una distribución predeterminada. Explícitamente, usando las expresiones (10) y (12), la complejidad se puede expresar de la siguiente manera:

$$C[P] = Q[P] \cdot H[P] \tag{13}$$

En este punto es importante destacar que esta definición de complejidad no requiere hacer ninguna suposición respecto del sistema que genera la señal en estudio, solamente se basa en la estructura de ordenamiento temporal contenida en la expresión (7). Por lo tanto, esta medida de complejidad estadística permite identificar diferentes niveles de periodicidad y caos (Zunino *et al.*, 2012). Dado que ambas medidas de información se calculan en términos de una FDP, en este trabajo se utiliza el enfoque propuesto por Bandt y Pompe (2002). Esta metodología consiste en la obtención de las probabilidades

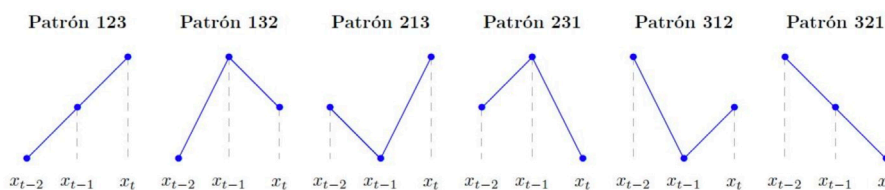


Fig. 2. Esquemización de Patrones ordinales para una dimensión de embedding $m = 3$

relacionadas con las permutaciones de los patrones ordinales (ver Figura 2) formadas por la secuencia de puntos que componen la señal en estudio.

Los métodos de estudio de series temporales basados en patrones ordinales (PO) (Bandt y Pompe, 2002) buscan cuantificar características de una señal midiendo la distribución secuenciada de un subconjunto de valores de la misma. En general, se estudia el orden de los patrones a partir de la posición relativa de los valores observados en subsegmentos de datos de longitud fija. De esta manera, se diferencian de los métodos tradicionales de análisis de series temporales no lineales, que en términos generales calculan parámetros en el espacio de fases reconstruido (Bradley y Kantz, 2015), o en aquellos que se obtienen basados en la dinámica simbólica.

A partir del trabajo germinal de Bandt y Pompe, gran parte de la literatura relativa al análisis ordinal aplicado se ha centrado en la investigación de señales medidas en sistemas biológicos complejos, Economía, y en diversidad de aplicaciones (cf. Amigó y Keller, 2015 y Zannin *et al*, 2012).

Las contribuciones más recientes en este campo incluyen el análisis ordinal de electrocardiograma (ECG) e intervalos entre latidos (Zannin *et al*, 2012, Zunino *et al*, 2015, Kulp *et al*, 2016), y el análisis del electroencefalograma (EEG) para la detección del inicio de convulsiones, la estadiación del sueño y la anestesiología (Keller *et al*, 2014, Zunino *et al*, 2015). La amplia diversidad de aplicaciones del método de PO se basa en que no es necesario suponer ninguna característica del sistema que produce la serie temporal, técnicamente se dice que es agnóstica al dominio (Cardozo Pereyra *et al*, 2022).

Hay dos aspectos para destacar en la teoría de los PO. Por un lado, bajo ciertas suposiciones existe una relación entre el espacio de fases y la dinámica simbólica de los PO (Keller y Sinn, 2010, Amigó y Keller, 2013). Por otro lado, se ha demostrado con respecto a los mapas discretos, que la distribución de símbolos ordinales enumerados a partir de una serie temporal se rige por las propiedades dinámicas del sistema subyacente (Bandt y Pompe, 2002, Amigó y Keller, 2013). Entre las ventajas de esta metodología se puede mencionar: la simplicidad conceptual y la sencillez de implementación computacional, la velocidad, la robustez y desde un punto de vista más teórico la invariabilidad con respecto a las transformaciones monotónicas no lineales de los datos originales (Amigó *et al*, 2007).

La gran mayoría de la literatura esencialmente se desarrolla en torno a cuantificar algún aspecto de la distribución estacionaria de los PO. Si bien es cierto que cada PO contiene información sobre la estructura temporal de corto alcance de las observaciones contiguas de una serie temporal (Bandt y Pompe, 2002), las cuestiones relativas a la relación entre la secuencia de patrones codificados en la dinámica simbólica ordinal y el comportamiento característico del sistema subyacente siguen siendo en gran medida inexploradas.

Sin embargo, recientemente se ha desarrollado una nueva medida denominada la entropía condicional de los PO que fue introducida por Unakafov y colaboradores (2015) para cuantificar la complejidad transitoria de la dinámica simbólica ordinal mediante el cálculo de la entropía para patrones sucesivos (Unakafov y Keller, 2015). Recientemente, en forma adicional, se ha propuesto incorporar el concepto de redes ordinales que ha sido complementado con el de redes complejas (Reik *et al*, 2011, Anselmo Cantú Ros *et al*, 2013, McCulloough *et al*, 2015), que a su vez permite cuantificar aspectos de la estructura temporal de la dinámica simbólica ordinal utilizando medidas sencillas del campo de la teoría de las redes.

Dado que los PO se calculan sobre segmentos de longitud fija de las señales y estas se codifican mediante un alfabeto preestablecido, se puede obtener la frecuencia con la que un segmento es seguido por otro de manera tal de recorrer toda la longitud permitida de la serie de datos. A su vez, estas frecuencias relativas se pueden emplear para aproximar

la probabilidad de que un segmento dado se comunique con otro presente en la señal. Esto constituye lo que se ha denominado probabilidad de transición entre PO (Zanin y Olivares, 2021).

Con todas las probabilidades de transición estimadas, se puede construir una matriz cuyas columnas contengan la probabilidad de transición de un PO dado hacia todos los otros posibles. La matriz así conformada será una matriz de probabilidad dado que la suma de los elementos de sus columnas dará como resultado 1 (para más detalles cf. Cardozo Pereira *et al.*, 2022).

Con los de los registros de monitoreo no intrusivo se construyeron las matrices de transición entre PO (MTPO) que posteriormente se emplearon para conformar el conjunto de características empleado por el algoritmo de aprendizaje automático.

Por último, se incluyó en el espacio de características para el algoritmo de aprendizaje automático la cantidad de cruces por el valor medio (MVCR) de cada una de las formas de onda de cada uno de los electrodomésticos. Este valor es simplemente sustraer la componente de continua de cada señal y contar cuantos cruces por cero presenta la misma.

Procesamiento

La cantidad de muestras que presentaba la base de datos para cada clase en un principio era reducida, para poder ampliar el número de muestras se subdividieron las señales originales. Esto fue posible porque cada electrodoméstico fue monitoreado por un intervalo de aproximadamente 2 meses, donde se repetían sus ciclos de funcionamiento.

Se utilizaron los siguientes criterios para la subdivisión de las señales originales: se fijó un nivel de consumo mínimo de 50 Watts y una longitud mínima de 50 segundos para todas las clases a excepción de las heladeras, clase para la cual la duración mínima, del ciclo de operación, seleccionada fue de 500 segundos, dado que resultó ser el electrodoméstico con el ciclo de operación más prolongado.

Los electrodomésticos de la base de datos original fueron: Aire acondicionado, Tomacorrientes Baño, Lavavajilla, Estufa Eléctrica, Electrónica, Calefactor, Tomacorriente cocina, Iluminación, Tablero principal, Microondas, Tomacorrientes desconocido, Horno, Heladera, Anafe, Tablero terminal, Lava secarropa, Triturador, Alarmas de incendio, Otros y Tomacorrientes de exterior.

Tras haber hecho pruebas preliminares con el objeto de simplificar los análisis de los resultados y posteriores interpretaciones, se redujeron las categorías de electrodomésticos a las siguientes clases:

- Aire Acondicionado
- Calefactor
- Microondas
- Horno
- Heladera
- Anafe

Para analizar la intensidad de la superposición entre las características se empleó su representación en el plano de entropía – complejidad ($H \times C$), tal como se puede observar en la Figura 3 hay un alto grado de superposición entre las características de naturaleza informacional, motivo por el cual se recurrió a la búsqueda de un algoritmo de aprendizaje automático que permitiese ampliar la diferenciación entre las clases de interés.

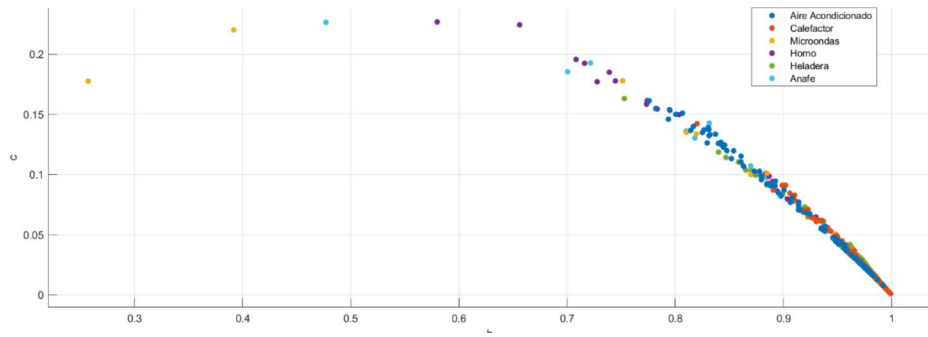
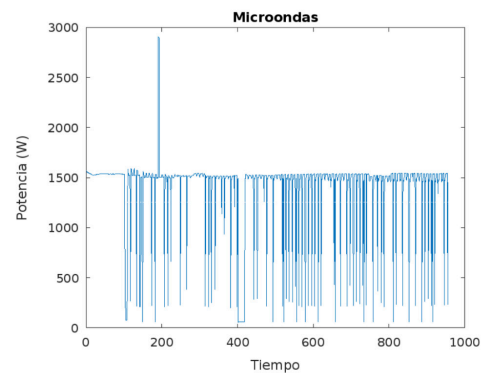
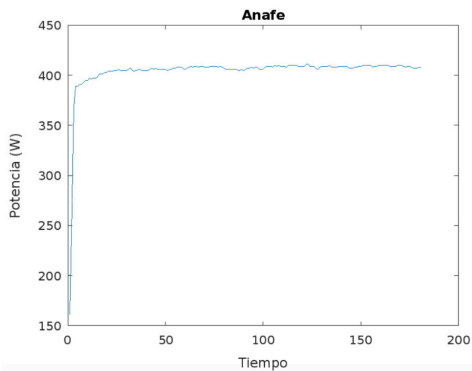
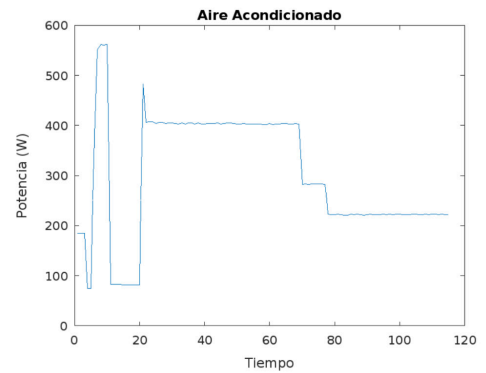
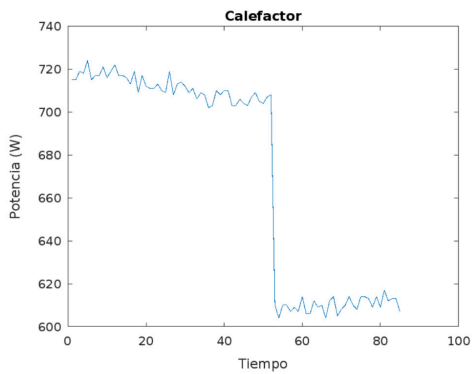


Fig. 3. Plano HxC donde se aprecia el alto grado de superposición entre clases

Tabla 1. Esquema del espacio de características

Objetivo (y_i)		Características (x_i)							
Casa	Electrodoméstico	Probabilidad de transiciones entre patrones ordinales				Herramientas informacionales			Otras
		123 - 123	321 - 123	EP	H_{Wn}	C	MVCR
1	Calefactor						
1	Anafe						
1
2	Microondas						
2	Anafe						
...
6	Aire Acondicionado						
					



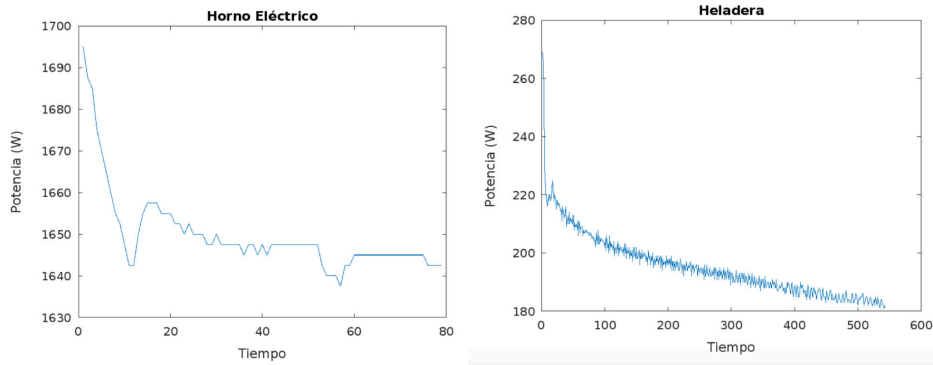


Fig. 4. Ejemplo de la forma de onda de las distintas señales

Al terminar el procesamiento de los datos, las dimensiones del conjunto resultante fueron de 833 registros con 40 dimensiones (36 correspondientes a las transiciones entre patrones, tres medidas informacionales y el MVCR) para las seis categorías en consideración, lo que supone un gran aumento comparado con la base original que contaba con un registro (o a lo sumo dos) por electrodoméstico, por cada una de las viviendas. En la Tabla 1 se presenta un esquema del espacio de características.

A continuación, se presentan en la Figura 4, gráficos de ejemplos de potencia en función del tiempo, correspondientes a los ciclos de funcionamiento para las clases seleccionadas en el presente estudio.

Máquina de soporte vectorial

Una máquina de soporte vectorial (SVM - *Support Vector Machines*) es un clasificador lineal binario que podrá distinguir las clases rotuladas como Ω_1 y Ω_2 a través de las siguientes relaciones:

$$y_i \in \Omega_1 \rightarrow w \cdot x_i + b \geq 1, \quad (14)$$

$$y_i \in \Omega_2 \rightarrow w \cdot x_i - b \leq -1, \quad (15)$$

donde

los parámetros del modelo serán $w \in \mathbb{R}^{s_v \times n}$ y $b \in \mathbb{R}^1$.

s_v es la cantidad de vectores de soporte,

mientras que el par (x_i, y_i) corresponde al i -ésimo registro x_i correspondiente al vector de características $e y_i \in \{\Omega_2: -1; \Omega_1: 1\}$ (Hearst et al, 1998. Bishop y Nasrabadi, 2006).

Maximización del margen

Dado un conjunto de muestras linealmente separables, existen diversas fronteras de decisión lineal tal como se muestra en la Figura 5.

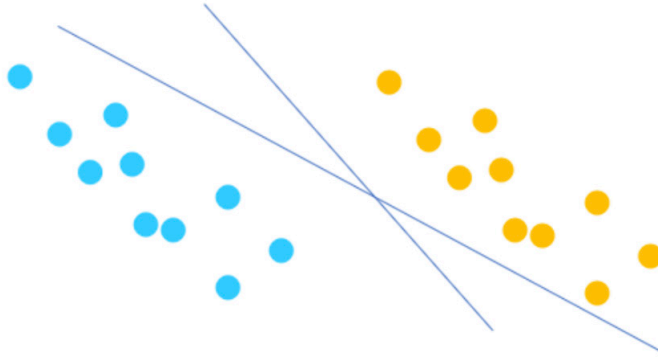


Fig. 5. Posibles clasificadores lineales que separan dos clases linealmente separables

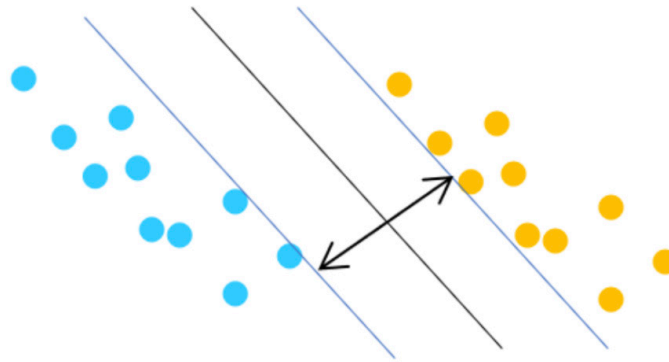


Fig. 6. Clasificador lineal con máximo margen de separación

Se deberá optar por algún criterio para la selección de esta frontera. En el caso del SVM se utilizará el concepto de maximización del margen entre clases, tal como se muestra en la Figura 6.

Para simplificar la notación entre clases se utilizará el valor de $y_i=1$ para indicar que $x_i \in \Omega_1$ de forma análoga se usará $y_i=-1$ para indicar que $x_i \in \Omega_2$

El ancho del margen quedará definido por la distancia entre los vectores soportes entre ambas clases. Si $x_1 \in \Omega_1$ y $x_2 \in \Omega_2$ son los vectores soportes, se pueden establecer las siguientes expresiones.

$$w \cdot x_1 + b = 1, \quad (16)$$

$$w \cdot x_2 + b = -1. \quad (17)$$

Que al restar (16) y (17) miembro a miembro y normalizando se obtiene el valor del margen ρ :

$$\rho: \frac{w}{|w|} (x_1 - x_2) = \frac{2}{|w|}. \quad (18)$$

Para maximizar el margen se debería plantear $2/|w|$, pero desde el punto de vista computacional es conveniente minimizar el recíproco y elevar la norma de w al cuadrado para que esta función objetivo sea una función cuadrática que resulte más simple de resolver:

$$\max\left(\frac{2}{|w|}\right) \equiv \min\left(\frac{|w|}{2}\right) \equiv \min\left(\frac{|w|^2}{2}\right). \quad (19)$$

En este problema de optimización también se deberá tener presente que el conjunto de datos de entrenamiento sea bien clasificado. Lo cual va a conformar el siguiente problema con restricciones:

$$\min_{w,b} \frac{|w|^2}{2}, \quad (20)$$

sujeito a: $y_i(w \cdot x_i + b) \geq 1 \forall i \in \Omega$.

Admisión de outliers, caso linealmente no separable

Cabe destacar que la formulación presentada tiene una falencia correspondiente al caso de que el conjunto de entrenamiento no sea linealmente separable, como en el caso de la Figura 7. Ya que la restricción impuesta exige que todo el conjunto de entrenamiento sea perfectamente clasificado.

Para sortear este obstáculo y permitir una resolución del problema planteado, se añade un término de regularización que penalizará los datos de entrenamiento mal clasificados a un costo $C \cdot \zeta$.

$$\min_{w,b,\zeta} \frac{|w|^2}{2} + C \sum_{i=1}^n \zeta_i. \quad (21)$$

Sujeito a: $y_i(w \cdot x_i + b) \geq 1 - \zeta_i, \forall i \in \Omega,$

$$\zeta_i \geq 0.$$

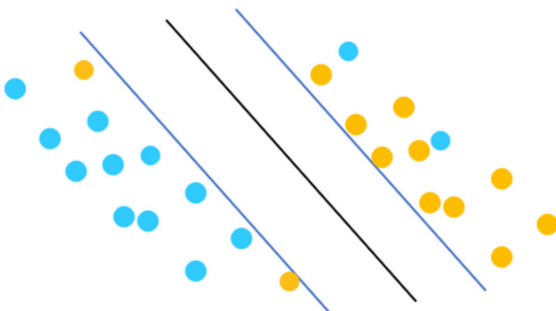


Fig. 7. La presencia de outliers imposibilita el uso de restricciones rígidas sin variables de holgura

La introducción de la variable de holgura ζ_i permitirá que los datos de entrenamiento puedan estar dentro del margen de separación, pero a un costo elevado definido por el parámetro C .

Problema dual de la maximización del margen y derivación de la función de decisión

Para la resolución de esta última formulación se planteará la función Lagrangiana para obtener un problema sin restricciones, donde las variables duales de las restricciones del problema de optimización (21) están representadas por α_i y β_i .

$$L(w, b, \zeta) = \frac{w^t w}{2} + C \sum_{i=1}^n \zeta_i + \sum_{i=1}^n \alpha_i (1 - \zeta - y_i (w^t x + b)) + \sum_{i=1}^n \beta_i (-\zeta_i). \quad (22)$$

De acuerdo con que esta función es convexa, el punto de ensilladura será el óptimo global del problema. Para encontrarlo se calcularán las derivadas con respecto a los parámetros y se igualarán a 0:

$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^n \alpha_i y_i \phi(x_i) = 0, \quad (23)$$

$$\frac{\partial L}{\partial b} = \sum_{i=1}^n \alpha_i y_i = 0, \quad (24)$$

$$\frac{\partial L}{\partial \zeta} = C - \alpha_i - \beta = 0, \quad i = 1..n. \quad (25)$$

Es interesante notar por las condiciones de Karush, Kuhn y Tucker (KKT) que en la expresión (23) el valor del multiplicador α_i será distinto de cero solo cuando la restricción esté activa. Como este multiplicador es mayor a cero ($\alpha_i \geq 0$) implica que si hay una restricción activa para algún punto x_j con $y_j=1$ deberá haber un punto x_k con $y_k=-1$ cuya restricción también se encuentre activa.

Reemplazando en la expresión de la función Lagrangiana por las expresiones (23), (24) y (25) se obtiene:

$$L(w, b, \zeta, \alpha) = \frac{w^t w}{2} + \sum_{i=1}^n \alpha_i - \sum_{i=1}^n \alpha_i y_i w^t x - \sum_{i=1}^n \alpha_i y_i b + C \sum_{i=1}^n \zeta_i - \sum_{i=1}^n \alpha_i \zeta_i - \sum_{i=1}^n \beta_i \zeta_i. \quad (26)$$

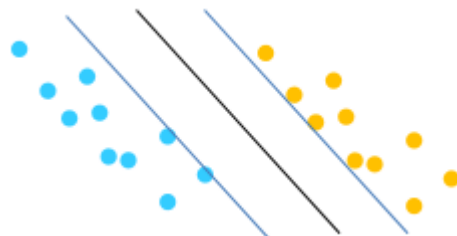


Fig. 8. Puntos donde están los vectores soporte, que coinciden con los puntos donde se encuentran los márgenes del clasificador

Los tres últimos términos suman cero de acuerdo con (25), luego el segundo término vale cero por (24), mientras que el primer término será la mitad del tercer sumando al sustituir w con lo obtenido en (23). Dando por resultado:

$$\max_{\alpha} L(\alpha) = \sum_{i=1}^n \alpha_i - \frac{\sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle}{2}. \quad (27)$$

Sujeto a: $\sum_{i=1}^n \alpha_i y_i = 0, \quad i = 1..n, \quad 0 \leq \alpha_i \leq C, \quad i = 1..n,$

que corresponde al modelo dual de optimización

donde:

solo intervienen los valores de α_i que serán distintos de cero si los puntos en la restricción $y_i(w \cdot x_i + b)$ del problema original es menor a uno (es decir o están sobre el margen o dentro del mismo). A este par de valores se los denomina vectores soporte (ver Figura 8) ya que en base a estos vectores y su variable dual asociada podrá crearse la función de decisión:

$$d(x; \alpha^*, b^*) = \text{sign} \left(\sum_{s_v} \alpha_i^* y_i \langle x, x_j \rangle + b^* \right). \quad (28)$$

Si el signo es positivo el punto x será clasificado como perteneciente a 1, en caso contrario se clasificará como perteneciente a Ω_2 . Cabe aclarar que la sumatoria indicada sólo tiene sentido sobre los vectores soporte (s_v) ya que para el resto de los puntos del dataset de entrenamiento su valor dual es nulo.

Extensión del SVM, uso del Kernel-Trick

Es interesante notar que la función de decisión dependerá de los valores de las variables duales del problema planteado anteriormente y de un producto escalar del nuevo dato que se quiere clasificar x y los vectores soporte x_j correspondientes al dataset de entrenamiento. Esto se puede interpretar como una medida de similitud de este nuevo dato x con respecto a los vectores soporte.

Para llevar este clasificador a que opere como un clasificador no lineal puede llevarse el dataset a un espacio de dimensión superior donde sea linealmente separable, como se ilustra en la Figura 9.

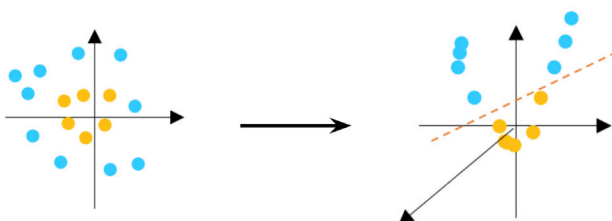


Fig. 9 Ejemplo de la aplicación del Kernel: los conjuntos que son difíciles de separar en el plano bidimensional pueden separarse de forma sencilla en el espacio tridimensional

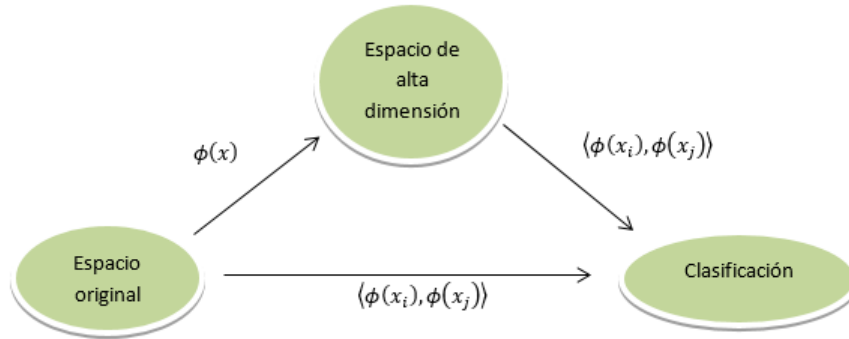


Fig. 10. Puntos donde están los vectores soporte, que coinciden con los puntos donde se encuentran los márgenes del clasificador

A través del teorema de Mercer (Bishop y Nasrabadi, 2006) se puede realizar esto de una manera simple. Se deberá escoger una función ϕ que sea una función de *kernel* cuya función de Gram (similitud par a par) sea definida positiva y utilizar este producto en el espacio de alta dimensión para obtener la clasificación. Que se simplifica a través de esta matriz de similitud (camino inferior en la Figura 10):

De esta manera la función de decisión resulta:

$$d(x; \alpha^*, b^*) = \underset{s_v}{\text{sign}} \left(\sum \alpha_i^* y_i \langle \phi(x), \phi(x_j) \rangle + b^* \right). \quad (29)$$

Algunas funciones de *kernel* utilizadas generalmente son:

<i>Kernel</i> Monómico	$\langle x, x' \rangle^q$
<i>Kernel</i> Polinómico	$(1 + \langle x, x' \rangle)^q$
Exponencial	$\exp\left(-\frac{ x-x' }{\gamma}\right)$
<i>Kernel</i> sobre función de base radial (Gaussiana)	$\exp\left(-\frac{ x-x' ^2}{\gamma}\right)$

También es destacable que, de acuerdo con las propiedades del producto escalar, se pueden tener una combinación lineal de diferentes funciones de *kernel*.

Resultados

En línea con la metodología de la sección anterior, se procede a exponer los resultados del clasificador. La clasificación se realizó utilizando la aplicación *Classification Learner de Matlab* con un modelo de SVM con un *kernel* Gaussiano cuya escala se ajustó automáticamente durante el entrenamiento, se utilizó una estandarización previa de los datos para realizar el entrenamiento y la generación de decisión multi-clase se realizó bajo el concepto de uno vs uno.

El método utilizado entrena al clasificador con un 80% de la base de datos y luego realiza una validación con el 20% restante. Comparando los resultados de dicha validación con los resultados reales se arma una matriz de confusión (Figura 11) en la que se observa que la clasificación de las categorías Heladera y Calefactor son las más efectivas, mientras que la categoría Anafé es la que peor se clasifica.

Matriz de Confusión

Valores Predichos	Aire Acondicionado	11 6.6%	2 1.2%	0 0.0%	0 0.0%	1 0.6%	0 0.0%	78.6% 21.4%
	Anafe	0 0.0%	1 0.6%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
	Calefactor	2 1.2%	0 0.0%	61 36.7%	2 1.2%	0 0.0%	0 0.0%	93.8% 6.2%
	Heladera	1 0.6%	0 0.0%	4 2.4%	74 44.6%	0 0.0%	0 0.0%	93.7% 6.3%
	Horno	0 0.0%	0 0.0%	0 0.0%	1 0.6%	4 2.4%	0 0.0%	80.0% 20.0%
	Microondas	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	2 1.2%	100% 0.0%
		78.6% 21.4%	33.3% 66.7%	93.8% 6.2%	96.1% 3.9%	80.0% 20.0%	100% 0.0%	92.2% 7.8%
	Valores Reales							
	Aire Acondicionado	Anafe	Calefactor	Heladera	Horno	Microondas		

Fig. 11. Matriz de confusión

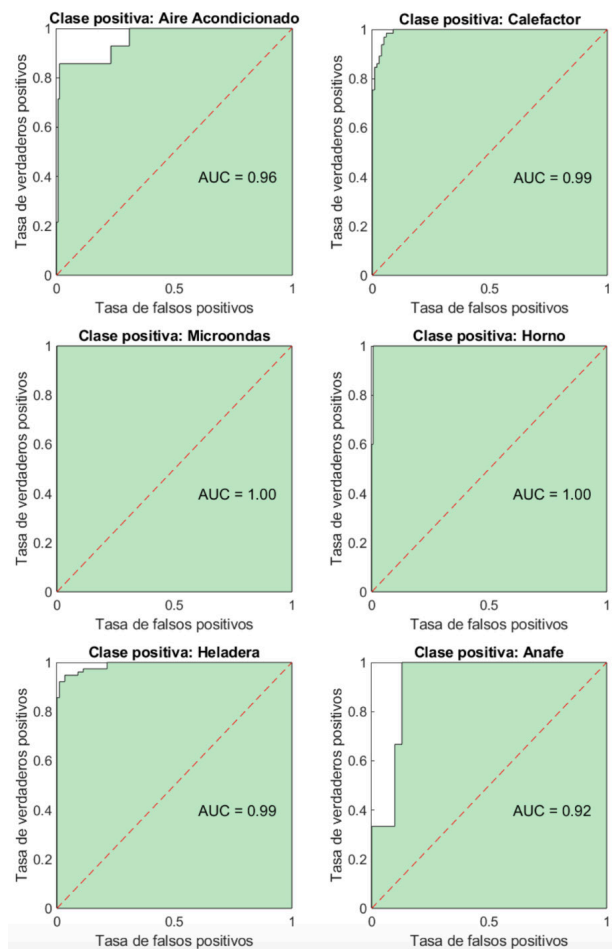


Fig. 12. Curvas ROC

Cabe destacar que las categorías Anafe, Microondas y Horno, comparadas con las otras, presentan una limitada cantidad de muestras y esto debe tenerse en cuenta para evitar sacar conclusiones erróneas sobre la efectividad del clasificador para estas categorías.

En la Figura 12 se observan las “curvas de característica operativa del receptor” o ROC, por sus siglas en inglés “*receiver operating characteristic*”, del modelo formulado correspondiente a cada categoría mediante la aplicación de SVM. Se puede destacar que el “área bajo las curvas ROC” o AUC, por sus siglas en inglés “*area under curve*”, es en todos los casos mayor a 92% lo cual indica un buen desempeño del clasificador.

En resumen, el modelo cuenta con una precisión de 86,1 %, un AUC siempre mayor a 92 % y una tasa general de verdaderos positivos de 92,2 %.

Conclusiones

En base a los resultados obtenidos, se puede concluir que la aplicación del algoritmo seleccionado de aprendizaje automático SVM sobre la forma de onda del consumo de electrodomésticos es una alternativa viable para elaborar modelos de clasificación que se puedan emplear en NILM, ya que el modelo utilizado obtuvo una tasa general de verdaderos positivos de 92,2 % y una precisión del 86,1 %.

Es importante entender las características de las categorías que se tuvieron en cuenta en el presente estudio y los resultados alcanzados para cada una de estas para poder mejorar la precisión del modelo en futuros trabajos y poder eventualmente aplicarlo en la sociedad.

En tal sentido, las heladeras y los calefactores son las categorías con la mayor cantidad de elementos analizados. Los mismos consiguieron con el modelo una tasa de verdaderos positivos de 96,1% y 93,8% respectivamente. Esto indica que, al menos para estas categorías en consideración, el modelo predice correctamente la gran mayoría de las veces.

Por el contrario, hay que resaltar que las categorías de Anafe, Horno y Microondas tienen una limitada cantidad de muestras. Una pequeña variación de predicciones acertadas de cualquiera de estas categorías modifica enormemente el porcentaje de la tasa de verdaderos positivos. Es por ello que no se puede afirmar el buen o mal desempeño del modelo para la detección de estos electrodomésticos. Es necesario ampliar la base de datos en futuros estudios que contemple un mayor número de artefactos de cada clase, como así también la incorporación del tipo de la tecnología asociada a los mismos, para mejorar la capacidad de clasificación de los modelos resultantes.

Agradecimientos

Los autores desean expresar su agradecimiento al programa de becas estudiantiles y de jóvenes graduados de la Universidad Tecnológica Nacional y de becas de estudiantes de la Secretaría de Asuntos Estudiantiles de la Facultad Regional Buenos Aires.

Referencias

- AMIGÓ J. M. y KELLER, K., (2013). Permutation entropy: One concept, two approaches. *The European Physical Journal Special Topics*, 222(2):263–273, 2013. DOI:10.1140/epjst/e2013-01840-1.
- AMIGÓ J. M.; KELLER, K. y UNAKAFOVA, V. A., (2015). Ordinal symbolic analysis and its application to biomedical recordings. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 373(2034):20140091, 2015. DOI:10.1098/rsta.2014.0091.
- AMIGÓ J. M.; ZAMBRANO, S. y SANJUÁN, M.A., (2007). True and false forbidden patterns in deterministic and random dynamics. *EPL (Europhysics Letters)*, 79(5):50001. DOI:10.1209/0295-5075/79/50001.
- AQUINO, A. L.; RAMOS, H. S.; FRERY, A. C.; VIANA, L. P.; CAVALCANTE, T. S. Y & ROSSO, O. A., (2017). Characterization of electric load with information theory quantifiers. *Physica A: Statistical Mechanics and its Applications*, 465, 277-284.
- BANDT, C. Y POMPE, B., (2002). Permutation entropy: a natural complexity measure for time series. *Physical Review Letters*, 88(17), 174102.
- BISHOP, C. M. y NASRABADI, N. M., (2006). *Pattern recognition and machine learning* (Vol. 4, No. 4, p. 738). New York: springer.
- BRADLEY, E. y KANTZ, H., (2015). Nonlinear time-series analysis revisited. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 25(9):097610. DOI:10.1063/1.4917289.
- BRISSAUD, J. B., (2005). The meanings of entropy. *Entropy*, 7(1), 68-96.
- CARDOSO-PEREIRA, I.; BORGES, J. B.; BARROS, P. H.; LOUREIRO, A. F.; ROSSO, O. A. y RAMOS, H. S., (2022). Leveraging the self-transition probability of ordinal patterns transition network for transportation mode identification based on GPS data. *Nonlinear Dynamics*, 107(1):889–908, 2022. DOI:10.1007/s11071-021-07059-x.
- DONNER, R. V.; SMALL, M.; DONGES, J. F.; MARWAN, N.; ZOU, Y.; XIANG, R. y KURTHS, J., (2011). Recurrence-based time series analysis by means of complex network methods. *International Journal of Bifurcation and Chaos*, 21(04):1019–1046. DOI:10.1142/S0218127411029021.
- FADLALLAH, B.; CHEN, B.; KEIL, A. y PRÍNCIPE, J., (2013). Weighted-permutation entropy: A complexity measure for time series incorporating amplitude information. *Physical Review E*, 87(2), 022911.
- FAWCETT, T., (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861-874.
- GARCIA CANTÚ ROS, A.; FORTI, G. y NICOLIS, G. (2013). Network representation of dynamical systems: Connectivity patterns, information and predictability. *Communications in Nonlinear Science and Numerical Simulation*, 18(8):2018–2034, 2013. DOI: 10.1016/j.cnsns.2012.11.007.
- HEARST, M. A.; DUMAIS, S. T.; OSUNA, E.; PLATT, J. y SCHOLKOPF, B., (1998). Support vector machines. *IEEE Intelligent Systems and their applications*, 13(4), 18-28.
- KELLER, K. y SINN, M., (2010). Kolmogorov–Sinai entropy from the ordinal viewpoint. *Physica D: Nonlinear Phenomena*, 239(12):997–1000. DOI:10.1016/j.physd.2010.02.006.
- KELLER, K.; UNAKAFOV, A. M. y UNAKAFOVA, V. A., (2014). Ordinal patterns, entropy, and EEG. *Entropy*, 16(12):6212–6239. DOI:10.3390/e16126212.
- KOLMOGOROV, A. N., (1958). New metric invariant of transitive dynamical systems and endomorphisms of Lebesgue spaces. In *Doklady of Russian Academy of Sciences* (Vol. 119, p. N5).
- KOLTER, J. Z. y JOHNSON, M. J., (2011, AUGUST). REDD: A public data set for energy disaggregation research. In *Workshop on data mining applications in sustainability (SIGKDD)*, San Diego, CA (Vol. 25, No. Citeseer, pp. 59-62).

- KULP, C. W.; CHOBOT, J. M.; FREITAS, H. R. y SPRECHINI, G. D., (2016). Using ordinal partition transition networks to analyze ECG data. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 26(7):073114, 2016. DOI:10.1063/1.4959537.
- LÓPEZ-RUIZ, R.; MANCINI, H. L. y CALBET, X., (1995). A statistical measure of complexity. *Physics Letters A*, 209(5-6), 321-326.
- MARTÍN, M. T.; PLASTINO, A. y ROSSO, O. A., (2006). Generalized statistical complexity measures: Geometrical and analytical properties. *Physica A: Statistical Mechanics and its Applications*, 369(2), 439-462.
- MCCULLOUGH, M.; SMALL, M.; STEMLER, T. y HO-CHING IU, H., (2015). Time lagged ordinal partition networks for capturing dynamics of continuous dynamical systems. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 142 25(5):053101. DOI:doi.org/10.1063/1.4919075.
- OLIVARES, F.; SOUZA, L.; LEGNANI, W. y ROSSO, O. A., (2019). Informational time causal planes: A tool for chaotic map dynamic visualization. In *Nonlinear Systems-Theoretical Aspects and Recent Applications*. IntechOpen.
- ROSSO, O. A.; DE MICCO, L.; LARRONDO, H. A.; MARTÍN, M. T. y PLASTINO, A., (2010). Generalized statistical complexity measure. *International Journal of Bifurcation and Chaos*, 20(03), 775-785.
- SANCHEZ GUEVARA, (2022). *British Petroleum statistical review of world energy, quantifying energy*. Energy News.
- SHANNON, C. E., (1948). A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review*, 5(1), 3-55.
- SINAI, I., (1959). On the concept of entropy for a dynamic system. *Doklady Akademii Nauk SSSR*, 124(4), 768-771.
- UNAKAFOV, A. M. y KELLER, K., (2014). Conditional entropy of ordinal patterns. *Physica D: Nonlinear Phenomena*, 144 269:94-102, 2014. DOI: 10.1016/j.physd.2013.11.015.
- ZANIN, M. y OLIVARES, F. (2021). Ordinal patterns-based methodologies for distinguishing chaos from noise in discrete time series. *Communications Physics*, 4(1), 1-14.
- ZUNINO, L.; OLIVARES, F. y ROSSO, O. A., (2015). Permutation min-entropy: An improved quantifier for unveiling subtle temporal correlations. *EPL (Europhysics Letters)*, 109(1):10005. DOI:10.1209/0295-5075/109/10005.
- ZUNINO, L.; SORIANO, M. C. y ROSSO, O. A., (2012). Distinguishing chaotic and stochastic dynamics from time series by using a multiscale symbolic approach. *Physical Review E*, 86(4), 046210.