1 Motivación: detección de la arritmia fibrilación auricular

Motivación: evaluación de la incertidumbre de pronósticos

- Construcción de modelos orientados por datos
- Modelo de producción de potencia solar FV normalizada
- Verosimilitud para el error de pronóstico
- Verosimilitud aproximada para el error de pronóstico

Proyecto ANII FSDA_1_2018_1_154651 - Creación de algoritmos utilizando técnicas de clasificación supervisada y no supervisada para el diagnóstico de enfermedades cardiovasculares en una población de adultos mayores de bajos recursos en Uruguay.

- Conjunto de registros electrocardiográficos (ECG). Dispositivo móvil de tecnología electrónica (DMTE) utilizado: sensor KardiaMobile de AliveCor vinculado con una aplicación Kardia (Alivecor inc.) que se instala en smartphones.
- El sistema detecta una única derivación del ECG realizando registros de 30 segundos de duración y genera tres posibles diagnósticos automáticos, que son: normal, posible fibrilación auricular (FA) o sin clasificar.

Descripción de los datos

- estudio "piloto" para la validación del dispositivo: 114 registros de ECG de adultos mayores de Uruguay (se dispone tanto de la etiqueta asignada por el DMTE como del diagnóstico realizado por expertos cardiólogos a partir del ECG de doce derivaciones, considerado el patrón oro para la detección de arritmias cardíacas);
- banco público de la competencia 2017 "PhysioNet/Computing in Cardiology (CinC)": 3332 registros de ECG de una sola derivación de duración 30 segundos (frecuencia de muestreo 300 Hz) que seleccionamos a partir del banco público de entrenamiento internacional de 8528 registros (con duración de 9 a 60 segundos) constituido para la competencia 2017 "PhysioNet/Computing in Cardiology (CinC)" (https://physionet.org/content/challenge-2017/1.0.0/)

Descripción de los datos

 estudio de detección de FA a nivel país en ambiente no controlado utilizando el dispositivo ya validado por el equipo investigador: colaboración entre la Comisión Honoraria para la Salud Cardiovascular (CHSCV) y el Plan Ibirapitá posibilitó la realización de talleres en todo el país, en los cuales, entre otras actividades, centenares de personas participaron del estudio de detección de FA, que consistió en que cada persona apoyara simultáneamente dos dedos de la mano derecha y dos dedos de la mano izquierda sobre los dos electrodos del DMTE por un período de tiempo suficiente para disponer de un registro de 30 segundos de duración.

La información recabada a través de los talleres, ha sido analizada por parte de un equipo de cardiólogas quienes, para cada registro de ECG que les fue asignado, llenaron un formulario electrónico elaborado exclusivamente para el proyecto. Cabe destacar que el procedimiento de asignación de los registros de

4/29

ECG al equipo de cardiólogas se realizó en base a un método de calibración estadística, diseñado de manera que la asignación de la etiqueta presencia/ausencia de FA a cada registro por parte de las cardiólogas no se viera influenciada por la posible variabilidad en el criterio de asignación entre las integrantes del equipo o por parte de una misma cardióloga.

Resultado: una base nacional de 423 registros de ECG que refiere expresamente a los adultos mayores, en la cual a cada registro se asocia la etiqueta FA o No FA.

= nar

イロト イポト イヨト イヨト



Salida original PDF del DMTE: parte del registro (110305).



◆□ > ◆□ > ◆ □ > ● □ > ◆ □ > ● □ > ◆ □ > ● □ >

Cuadro superior: registro (105629) de ECG clasificado como normal por el DMTE y los expertos. Cuadro inferior: registro (110305) de ECG clasificado como FA por el dispositivo y los expertos.



30 de agosto de 2022 8/29

Cuadro superior: registro (113323) de ECG no clasificado por el dispositivo y clasificado como NO FA por los expertos. Cuadro inferior: registro (110801) de ECG no clasificado por el dispositivo y clasificado como FA por los expertos.



Extracción de un conjunto de características de los ECG basadas en los intervalos RR y sus primeras diferencias; creación de nuevas características:

- media de los intervalos RR que tienen una duración mayor al percentil 90;
- media de los intervalos RR comprendidos entre el percentil 50 y el percentil 80;
- la diferencia de las dos medias anteriores.

Entrenamiento de batería de algoritmos de aprendizaje estadístico utilizando un conjunto de 3332 registros de la competencia Physionet 2017 PhysioNet/CinC con 19 características seleccionadas para cada registro.

Valoración de la capacidad predictiva de cada algoritmo emplenado un esquema de entrenamiento de tipo validación cruzada 10-fold repetido 5 veces.

Resultados para la base nacional de 423 registros de ECG: se encontraron algoritmos de mejor rendimiento para el problema en estudio:

- minimizan los falsos negativos, o sea que un verdadero registro de FA sea clasificado como Normal;
- identifican correctamente todos los registros informados como FA por los cardiólogos, presentando leves variaciones en la cantidad de falsos positivos;
- comparando el rendimiento de estos algoritmos con los resultados automáticos de la aplicación del DMTE, se observa que los nuevos algoritmos generados logran la correcta clasificación de la mayoría de los registros "Sin clasificar", donde dos de ellos clasifican incorrectamente sólo 7 de los 58 registros "Sin clasificar".

(ロ) (部) (三) (三) (三) (1/29)

Descripción de los datos

Observaciones de producción diaria de energía solar fotovoltaica (cada diez minutos) y pronósticos a un día (cada hora) en Uruguay durante 2019, normalizados con respecto a la capacidad máxima de energía solar instalada de 228.8 MW.



Figura 1: Producción de potencia solar FV en Uruguay (línea azul), pronóstico a un día (línea verde), máxima potencia estimada (línea anaranjada).

Uruguay - plantas solares

PROYECTOS OPERA	TIVOS			Actualización	agosto 2019
Convocatoria	Empresa	Nombre de la central	Potencia (MW)	Departamento	Entrada en operación
-	MIEM/UTE	Asahi	0,48	Salto	Set-2013
133/013	FRV S.A.	La Jacinta	50	Salto	Set-2015
R 131317	Raditon S.A.	Parque I. de Paysandú	8	Paysandú	Dec-2015
133/013	Alto Cielo S.A.	Alto Cielo	20	Artigas	Mar-2016
133/013	YARNEL S.A.		9,5	Río Negro	Jun-2017
R 131318	Casalko S.A.	Casalko	1,75	Paysandú	feb-17
133/013	JOLYPARK S.A.	Del Litoral	16	Salto	jun-17
133/013	NATELU S.A.		9,5	Soriano	jul-17
133/013	Giacote S.A.	Giacote Arapey	10	Artigas	ago-17
133/013	Giacote S.A.	Giacote Menafra	20	Río Negro	ago-17
R 131319	Petilcoran		9,5	Paysandú	ago-17
R 131320	Fenima S.A.		9,5	Paysandú	ago-17
R 131316	Dicano S.A.		11,25	Paysandú	ago-17
133/013	COLIDIM S.A.	El Naranjal	50	Salto	sep-17
133/013	Vingano S.A.		1	Paysandú	jun-18
133/013	Edelbon S.A.		1	Paysandú	jun-18
133/013	Lafemir S.A.		1	Paysandú	jun-18
	MIEM/UTE	Hikari	0,25	Lavalleja	jun-19
	Total		228,73		

Modelos orientados por datos

Sea $X = \{X_t \in [0, h_t], t \in [0, T]\}$ el proceso estocástico que modela la producción de potencia solar FV normalizada, definida por la siguiente ecuación diferencial estocástica (EDS) de Itô:

 $\begin{cases} dX_t = a(X_t; p_t, \dot{p}_t, \theta) dt + b(X_t; h_t, p_t, \dot{p}_t, \theta) dW_t, & t \in [0, T] \\ X_0 = x_0 \in [0, h_0] \end{cases}$ (1)

donde:

- $a(\cdot, p_t, \dot{p}_t, \theta) : [0, 1] \rightarrow \mathbb{R}$ indica una función de deriva,
- $b(\cdot; h_t, p_t, \dot{p}_t, \theta) : [0, 1] \rightarrow \mathbb{R}_+$ es una función de difusión,
- θ es un vector de parámetros desconocidos,
- $(p_t)_{t \in [0,T]}$ es el pronóstico disponible, que toma valores en $[0, h_t]$ y $(\dot{p}_t)_{t \in [0,T]}$ es su derivada respecto al tiempo,
- (*h*_t)_{t∈[0,T]} es la cota superior estimable para el fenómeno observado y su pronóstico,
- $(W_t)_{t \in [0,T]}$ es un proceso de Wiener estándar a valores reales.

14/29

Modelo de producción de potencia solar FV normalizada

Modelo de producción de potencia solar FV

$$\begin{cases} dX_t = (\dot{p}_t - \theta_t (X_t - p_t))dt + \sqrt{2\alpha\theta_0 X_t (h_t - X_t)}dW_t, & t \in [0, T] \\ X_0 = x_0 \in [0, h_0] \end{cases}$$

Cambio de variable $Y_t = \frac{X_t}{h_t}$

(2

Modelo de producción de potencia solar FV normalizada

Modelo de producción de potencia solar FV normalizada

$$\begin{cases} dY_t = \left(\left(\frac{\dot{p}}{h}\right)_t - \left(\theta_t + \frac{\dot{h}t}{h_t}\right)(Y_t - \frac{p_t}{h_t})\right) dt + \sqrt{2\alpha\theta_0 Y_t(1 - Y_t)} dW_t, t \in [\tilde{t}, \tilde{T}] \\ Y_{\tilde{t}} = y_{\tilde{t}} \in [0, 1] \end{cases}$$
(3)
donde $\left(\frac{\dot{p}}{h}\right)_t = \frac{\dot{p}_t h_t - \dot{h}_t p_t}{h_t^2}.$

[Teorema (Existencia y unicidad)] Sea $\tau_0 := \inf\{t \in [\tilde{t}, \tilde{T}], Y_t = 0\}$ y $\tau_1 := \inf\{t \in [\tilde{t}, \tilde{T}], Y_t = 1\}$ con $\inf \emptyset = +\infty$. Supongamos que

$$\forall t \in [\tilde{t}, \tilde{T}], \ 0 \leq \left(\frac{\dot{p}}{h}\right)_t + \left(\theta_t + \frac{\dot{h}t}{h_t}\right) \frac{p_t}{h_t} \leq \theta_t + \frac{\dot{h}t}{h_t} \leq C_1 < \infty.$$
 (A)

Entonces existe una única solución fuerte a (3) tal que para todos $t \in [\tilde{t}, \tilde{T}]$, $Y_t \in [0, 1]$.

Además, si supongamos que $y_{\tilde{t}} \in]0, 1[, \frac{p_t}{h_t} \in]0, 1[$ para todos $t \in [\tilde{t}, \tilde{T}]$, y que se fortalece la condición (A)

$$0 < \alpha \theta_0 \le \left(\frac{\dot{p}}{h}\right)_t + \left(\theta_t + \frac{\dot{h}t}{h_t}\right) \frac{p_t}{h_t} \le \left(\theta_t + \frac{\dot{h}t}{h_t}\right) - \alpha \theta_0 \le C_1 < \infty.$$
 (A')

Entonces, $\tau_0 = \tau_1 = +\infty$ c.s.

Por lo tanto, dado $y_{\tilde{t}} \in]0, 1[$, la única solución fuerte de (3) tendrá trayectorias casi seguramente en]0, 1[si el parámetro de reversión a la media variable en el tiempo $\left(\theta_t + \frac{ht}{h_t}\right)$ para el proceso Y_t cumple la siguiente condición:

$$\theta_t + \frac{\dot{h}t}{h_t} \ge \max\left(\frac{\alpha\theta_0 + \left(\frac{\dot{p}}{h}\right)_t}{1 - \frac{p_t}{h_t}}, \frac{\alpha\theta_0 - \left(\frac{\dot{p}}{h}\right)_t}{\frac{p_t}{h_t}}\right), \ \forall t \in [\tilde{t}, \tilde{T}].$$
(B)

Obs. La condición (B) muestra que, dada una cota superior suave h_t y un pronóstico suave p_t , en general no es posible hallar un parámetro variable en el tiempo acotado θ_t , porque la cota inferior puede llegar a ser no acotada cuando $\frac{p_t}{h_t} \rightarrow 0$ o $\frac{p_t}{h_t} \rightarrow 1$.

Función de pronóstico truncada

 Nuestro enfoque: se considera la siguiente función de pronóstico truncada

$$\left(\frac{p_t}{h_t}\right)^{\epsilon} = \begin{cases} \epsilon & \text{if } \frac{p_t}{h_t} < \epsilon \\ \frac{p_t}{h_t} & \text{if } \epsilon \le \frac{p_t}{h_t} < 1 - \epsilon \\ 1 - \epsilon & \text{if } \frac{p_t}{h_t} \ge 1 - \epsilon \end{cases}$$
(4)

que cumple $\left(\frac{p_t}{h_t}\right)^{\epsilon} \in [\epsilon, 1 - \epsilon]$ para cada $0 < \epsilon \ll \frac{1}{2}$ y $t \in [\tilde{t}, \tilde{T}]$, de manera que θ_t es acotado para cada $t \in [\tilde{t}, \tilde{T}]$. En la práctica, se debe lograr la calibración del parámetro de umbral ϵ . En este trabajo, se considera ϵ como un parámetro adicional del modelo que, junto con $\theta = (\theta_0, \alpha)$, puede ser estimado empleando técnicas basadas en la verosimilitud.

(ロ) (四) (三) (三) (三) (三) (19/29)

Error de pronóstico de la producción de potencia solar FV normalizada

El cambio de variable $V_t = Y_t - \frac{p_t}{h_t}$ permite plantear

Modelo para el error de pronóstico de la producción de potencia solar FV normalizada

$$\begin{cases} dV_t = -\left(\theta_t + \frac{\dot{h}_t}{h_t}\right) V_t dt + \sqrt{2\alpha\theta_0 \left(V_t + \frac{p_t}{h_t}\right) \left(1 - V_t - \frac{p_t}{h_t}\right)} dW_t, t \in [\tilde{t}, \tilde{T}] \\ V_{\tilde{t}} = v_{\tilde{t}} \in [-\frac{p_{\tilde{t}}}{h_t}, 1 - \frac{p_{\tilde{t}}}{h_t}]. \end{cases}$$

(5)

Verosimilitud para el error de pronóstico

- *M* trayectorias del proceso de Itô en tiempo continuo *V*.
- se observa cada trayectoria en N + 1 puntos con distancia Δ .
- se indica la muestra aleatoria con

$$V^{M,N+1} = \left\{ V_{t_1}^{N+1}, V_{t_2}^{N+1}, \dots, V_{t_M}^{N+1} \right\},$$

donde t_j es el tiempo inicial de la trayectoria j y $V_{t_j}^{N+1} = \left\{ V_{t_j+i\Delta}, i = 0, \dots, N \right\}, \forall j \in \{1, \dots, M\}.$

Sea $\rho(v|v_{j,i-1}; \theta, \epsilon)$ la densidad de probabilidad condicional de $V_{t_j+i\Delta} \equiv V_{j,i}$ dado $V_{j,i-1} = v_{j,i-1}$ evaluada en v, donde $\theta = (\theta_0, \alpha)$ son los parámetros desconocidos del modelo y ϵ es el umbral desconocido.

• El proceso de Itô definido por la EDS (5) es Markoviano.

Verosimilitud para el error de pronóstico

• La función de verosimilitud de la muestra V^{M,N+1} es:

$$\mathcal{L}\left(\boldsymbol{\theta},\varepsilon;\boldsymbol{V}^{M,N+1}\right) = \prod_{j=1}^{M} \left\{ \prod_{i=1}^{N} \boldsymbol{\rho}\left(\boldsymbol{V}_{j,i} | \boldsymbol{V}_{j,i-1}; \boldsymbol{\rho}_{[t_{j,i-1},t_{j,i}]}, \boldsymbol{\theta},\varepsilon\right) \right\}, \quad (6)$$

donde $t_{j,i} \equiv t_j + i\Delta$ para cada $j = 1, \dots, M$ y $i = 0, \dots, N$.

Técnica de igualdad de momentos

- Expresiones analíticas exactas para las densidades de transición de *V*, $\rho(V_{j,i}|V_{j,i-1}; \theta, \epsilon)$ son raramente disponibles.
- Métodos de verosimilitud aproximada
- Técnica de igualdad de momentos:
 - se considera una densidad de transición aproximada para V.
 - se igualan los momentos condicionales de la densidad aproximada para *V* con los momentos condicionales del modelo EDS (5).

Log-verosimilitud aproximada para el error de pronóstico

Para cada t ∈ [t_{j,i-1}, t_{j,i}], aproximamos las densidades de transición del proceso V usando una distribución beta con parámetros de forma ξ₁ y ξ₂. Igualando los primeros dos momentos centrales de V con los correspondientes momentos de la distribución beta proxy con dominio [-1 + ε, 1 - ε], los parámetros de forma están dados por

$$\xi_{1}(t) = -\frac{(\mu_{t} + 1 - \epsilon)(\mu_{t}^{2} + \sigma_{t}^{2} - (1 - \epsilon)^{2})}{2(1 - \epsilon)\sigma_{t}^{2}}$$

$$\xi_{2}(t) = \frac{(\mu_{t} - 1 + \epsilon)(\mu_{t}^{2} + \sigma_{t}^{2} - (1 - \epsilon)^{2})}{2(1 - \epsilon)\sigma_{t}^{2}},$$

donde $\mu_t = m_1(t)$ y $\sigma_t^2 = m_2(t) - m_1(t)^2$.

• La log-verosimilitud $\tilde{\ell}(\cdot; v^{M,N+1})$ de la muestra observada $v^{M,N+1}$:

Log-verosimilitud aproximada para el error de pronóstico

$$\tilde{\ell}(\theta,\epsilon; v^{M,N+1}) = \sum_{j=1}^{M} \sum_{i=1}^{N} \log \left\{ \frac{1}{2(1-\epsilon)} \frac{1}{B(\xi_{1}(t_{j,i}^{-}),\xi_{2}(t_{j,i}^{-}))} \left(\frac{v_{j,i}+1-\epsilon}{2(1-\epsilon)}\right)^{\xi_{1}(t_{j,i}^{-})-1} \times \left(\frac{1-\epsilon-v_{j,i}}{2(1-\epsilon)}\right)^{\xi_{2}(t_{j,i}^{-})-1} \right\}, \quad (8)$$

donde los parámetros de forma $\xi_1(t_{j,i}^-)$ y $\xi_2(t_{j,i}^-)$, según (7), dependen de $\mu(t_{j,i}^-; \theta)$ y $\sigma^2(t_{j,i}^-; \theta)$ cuando $t \uparrow t_{j,i}$, que calculamos resolviendo numéricamente el problema con valores iniciales para las ecuaciones de los momentos. $B(\xi_1, \xi_2)$ indica la función beta.

◆□ → < □ → < Ξ → < Ξ → Ξ </p>
⑦ < ♡ < 25/29</p>

Aplicación: datos solar FV y pronóstico en Uruguay



Figura 2: Conjuntos de nivel de la log-verosimilitud negativa como función de (θ_0, α) .

A B > A B > A B
 A

- ₹ ₹ >

Aplicación: datos solar FV y pronóstico en Uruguay



Figura 3: Cinco trayectorias simuladas de la producción de potencia solar FV normalizada con beta proxy, MLE (θ_0, α) = (23.30, 0.15) y ε = 0.07.

<ロ> <同> <同> < 同> < 同> < 同> <

Aplicación: datos solar FV y pronóstico en Uruguay



Figura 4: Bandas de confianza para la producción de potencia solar FV normalizada con beta proxy, MLE (θ_0, α) = (23.30, 0.15) y ε = 0.07.

¡Muchas gracias por su atención!

イロン イロン イヨン イヨン