Pontificia Universidad Católica de Chile
Escuela de Ingeniería
Departamento de Ingeniería Eléctrica

Reporte de Modelos

Nawel Carimán Fuenzalida

17 de Septiembre del 2020

---

1. Metodología

1.1. Procesamiento de datos

Base de datos

La base de datos utilizada cuenta con datos entre el 2020-01-24 17:00:00 al 2020-01-30 16:55:00, es decir, 6 días de información. La variable a predecir será la de medición continua de glucosa *sensor_glucose*. Este dato se encuentra en una tabla SQL junto a otras variables entregadas por el monitor continuo. Los datos fueron ordenados para tener una tasa de muestreo T_s de 5 minutos. Se utilizará la variable $y(t)$ para denotar la variable *sensor_glucose* medida en el tiempo t.

Pre procesamiento

La variable del sensor continuo de glucosa $y(t)$ cuenta con una pequeña pérdida de datos (5 valores), por lo que se realizó una interpolación lineal para manejar este problema. No se utilizó un mecanismo más sofisticado ya que la cantidad de información perdida es poca a comparación de la totalidad de los datos.

Datos de entrenamiento y prueba

El conjunto de datos se dividió en entrenamiento y prueba, con 4 y 2 días respectivamente, es decir, 66,7 % de entrenamiento y 33,3 % de prueba.

1.2. Entrenamiento, predicción y desempeño

Para cada algoritmo, se busca generar un modelo predictivo de un paso adelante (5 minutos). Luego, de modo recursivo, se obtendrá una trayectoria de predicción de seis pasos adelante (30 minutos) para finalmente evaluar el desempeño bajo distintos indicadores.

Denotaremos a la predicción como $\hat{y}_{t+k}(t), k = 1, \ldots, 6$ como la predicción de k pasos adelante al tiempo t de y.

Entrenamiento y predicción

Cada algoritmo de predicción será detallado en cada sección, donde se mostrará un gráfico de la variable $y(t)$ y las predicciones a futuro.

Desempeño

Los indicadores de desempeño se indican a continuación:

- Error de predicción: Para cada modelo, se calcularán tres tipos de errores:

  1. Error de un paso adelante: Este se define como

  $$\epsilon_{t+1}(t) = \hat{y}_{t+1}(t) - y(t) \tag{1}$$

  2. Error de seis paso adelante: Este se define como

  $$\epsilon_{t+6}(t) = \hat{y}_{t+6}(t) - y(t) \tag{2}$$

  3. Error de trayectoria: Este se define como

  $$\epsilon_{trajectory}(t) = \left[\frac{1}{6}\sum_{k=1}^{6}\left(\hat{y}_{t+k}(t) - y(t)\right)^2\right]^{1/2} \tag{3}$$

  Notar que el error de trayectoria cuenta una cota inferior (valor mínimo posible es cero) a diferencia de los demás errores. Luego, se presentará como resultado un resumen estadístico e histograma del error, gráficos en función del tiempo y un análisis en frecuencia, donde se mostrará el periodograma definido como

  $$Y_N(k) = \left|\frac{1}{\sqrt{N}}\sum_{t=1}^{N} y(t) e^{\frac{2\pi k i t}{N}}\right|^2 \tag{4}$$

  para $k = 1, \ldots, N$. También se mostrará una estimación del espectro, definida como

  $$\hat{\Phi}_y^N(\omega) = \sum_{\tau=-\gamma}^{\gamma} w_\gamma(\tau)\hat{R}_y^N(\tau)e^{-i\tau\omega} \tag{5}$$

  con $w_\gamma(\tau)$ una función ventana y $\hat{R}_y^N(\tau)$ la función de autocorrelación definida como

  $$\hat{R}_y^N(\tau) = \frac{1}{N} = \sum_{t=\tau}^{N} u(t)u(t-\tau) \tag{6}$$

  El valor de γ suele estar limitado a $\gamma = \pm N/2 - 1$, valor que se utilizará generalmente a menos que se indique lo contrario.

- Error cuadrático medio (RMSE): Este se define como:

  $$RMSE_i = \left[\frac{1}{N}\sum_{k=1}^{N}\epsilon_i(k)^2\right]^{1/2} \tag{7}$$

donde N es el número total de puntos y $\epsilon_i(k)$ son los tres errores descritos previamente, obteniendo tres errores cuadráticos medios; uno para un paso adelante, uno para seis pasos adelante y uno para la trayectoria.

- Ganancia temporal (TG): Esta se define como:

$$delay = \arg\min_{i\epsilon[0,L]} \left\{ \frac{1}{N-L} \sum_{k=1}^{N-L} \left(\hat{y}_{t+6}(k+i) - y(k)\right)^2 \right\} \tag{8}$$

$$TG = (L - delay) \cdot \Delta t \tag{9}$$

con Δt correspondiente al tiempo de muestreo y L el horizonte de predicción.

- Energía normalizada de la diferencia de segundo orden (ESOD-n): Esta se define como:

$$ESOD_n = \frac{ESOD(\hat{y}_{t+6})}{ESOD(y)} \tag{10}$$

$$= \frac{\sum_{k=3}^{N} \left(\hat{y}_{t+6}(k) - 2\hat{y}_{t+6}(k-1) + \hat{y}_{t+6}(k-2)\right)^2}{\sum_{k=3}^{N} \left(y(k) - 2y(k-1) + y(k-2)\right)^2} \tag{11}$$

2. Resultados - Resumen

En la tabla 1 se resume el RMSE para los distintos modelos entrenados hasta el momento.

	Entrenamiento			Prueba		
	ϵ_{t+1}	ϵ_{t+6}	$\epsilon_{trajectory}$	ϵ_{t+1}	ϵ_{t+6}	$\epsilon_{trajectory}$
Modelo de persistencia	6.95	34.88	23.53	6.98	33.17	22.66

Cuadro 1: Resumen del RMSE para los distintos modelos

3. Modelo de persistencia

3.1. Descripción

El primer modelo que se utilizará es el de persistencia. Este consiste en mantener el último valor y proyectarlo hacia el futuro. Por lo tanto, el modelo es

$$y(t) = y(t-1) + \epsilon_t \tag{12}$$

Y con esto, la predicción es

$$\hat{y}_{t+1}(t) = y(t-1) \tag{13}$$

3.2. Entrenamiento

Este modelo no requiere entrenamiento.

3.3. Resultados

En las figuras 1 y 2 se muestra un gráfico para la predicción de la glucosa para todas las predicciones y para la de seis pasos adelante respectivamente.

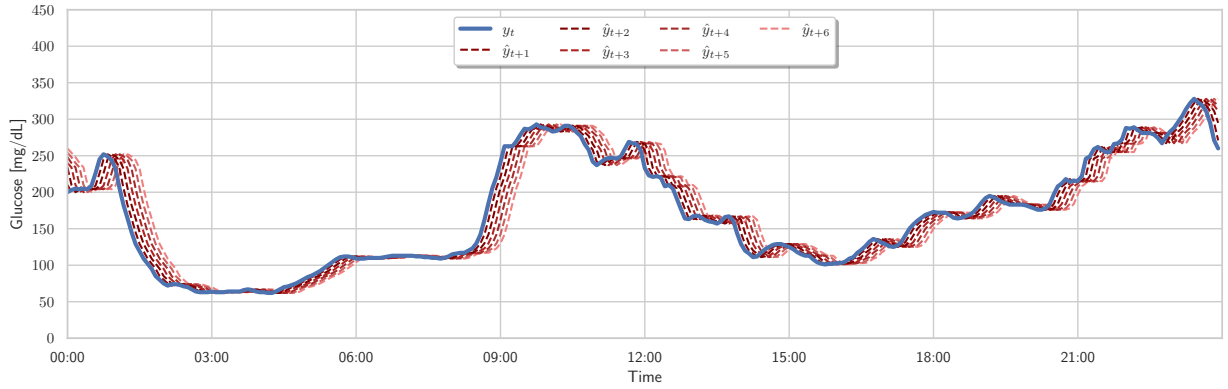

Figura 1: Gráfico de predicción de glucosa para todas las predicciones

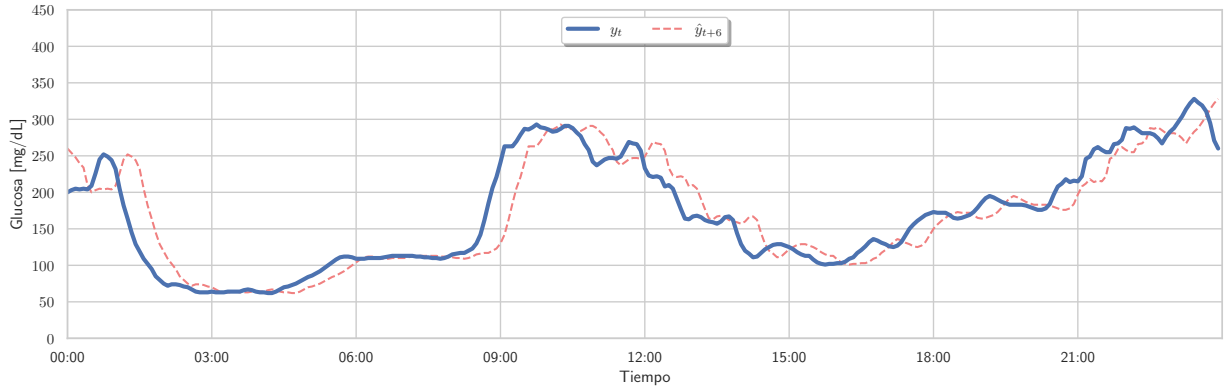

Figura 2: Gráfico de predicción de glucosa para seis pasos adelante

Error de predicción

En las figuras 3 y 4 se puede ver los distintos errores en función del tiempo para el conjunto de entrenamiento como de prueba. Podemos notar que si bien estamos utilizando el modelo más sencillo, hay periodos de tiempo donde el error es cercano a cero, mientras que en otros momentos existen errores superiores a 100 mg/dL, posiblemente explicados por por la ingesta de alimentos o acción de la insulina.

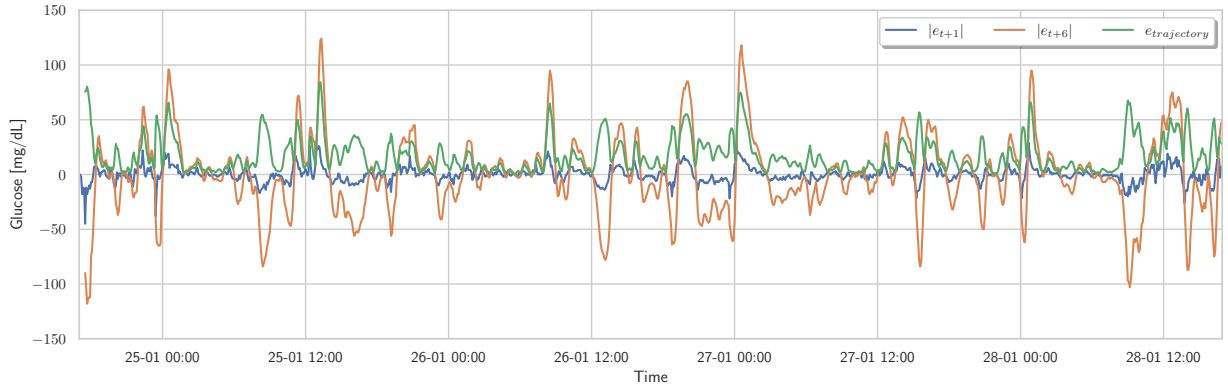

Figura 3: Gráfico del error en función del tiempo para el conjunto de entrenamiento

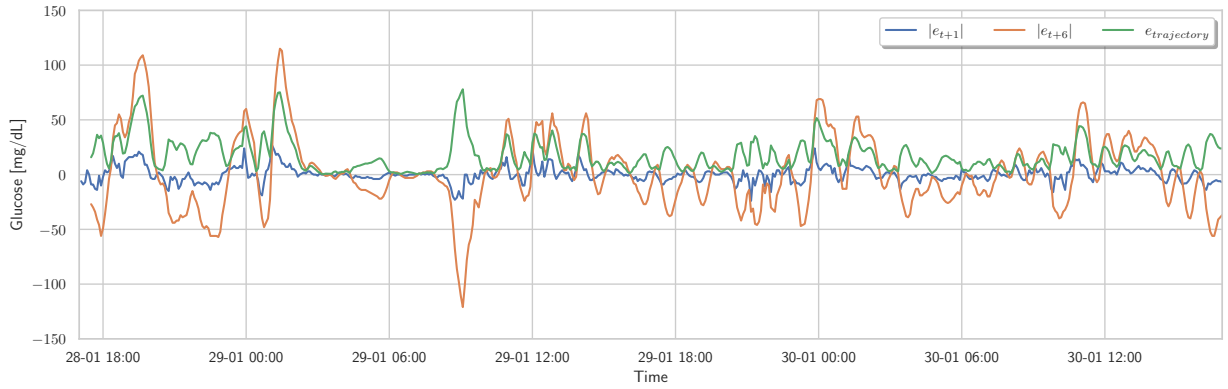

Figura 4: Gráfico del error en función del tiempo para el conjunto de prueba

En la tabla 2 se puede ver el resumen estadístico de los errores para el conjunto de entrenamiento como de prueba, mientras que en la figura 5 se muestran los histogramas para cada conjunto. Notar que la media y la mediana para ϵ_{t+1} y ϵ_{t+6} son cercanos a cero, y su histograma tiene forma gaussiana, lo que se condice con los gráficos de tiempo mostrados.

	Entrenamiento			Prueba		
	$\epsilon_{t+1}(t)$	$\epsilon_{t+6}(t)$	$\epsilon_{trajectory}(t)$	$\epsilon_{t+1}(t)$	$\epsilon_{t+6}(t)$	$\epsilon_{trajectory}(t)$
Número de datos	1151	1146	1146	575	570	570
Media	-0.091	-0.54	17.45	0.071	0.73	17.38
Desviación estándar	6.95	34.89	15.79	6.98	33.19	14.55
Mínimo	-45	-118	0.7	-24	-121	0.41
25 %	-4	-19	5.4	-4	-19	6.44
50 %	0	-1	11.95	0	-2	13.17
75 %	3	15	24.99	3	18	25.27
Máximo	29	124	84.31	27	115	77.91

Cuadro 2: Resumen estadístico del error

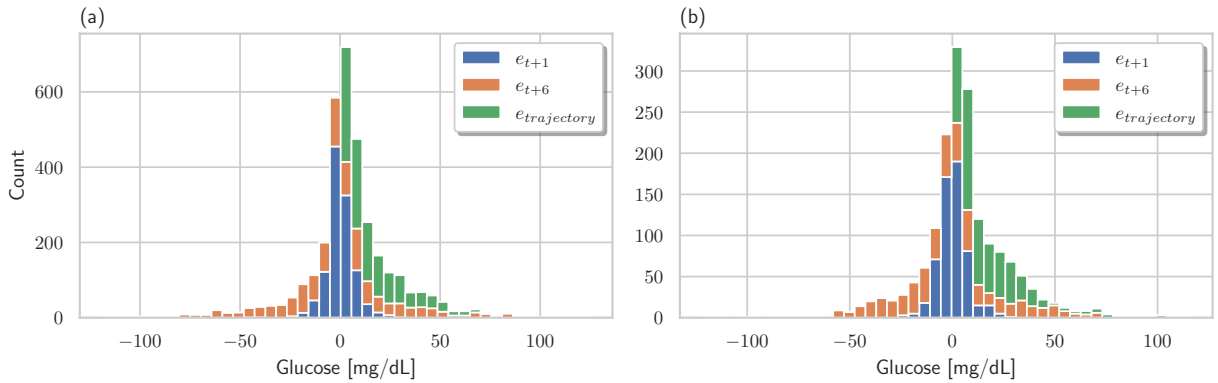

Figura 5: (a) Histograma para el conjunto de entrenamiento; (b) Histograma para el conjunto de prueba

En las figuras 6 y 7 se muestra el periodograma para cada conjunto, mientras que en las figuras 8 y 9 se muestra una estimación del espectro para una ventana hanning con $\gamma = N/2 - 1$.

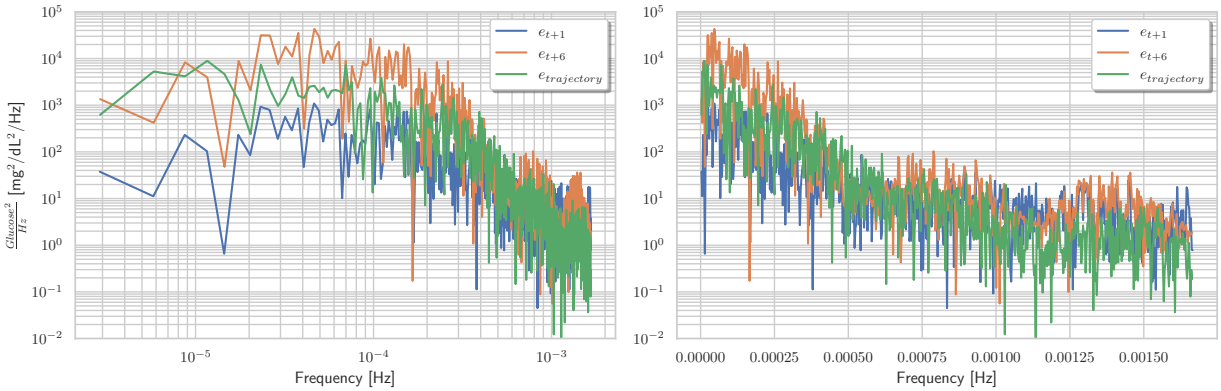

Figura 6: Gráfico del periodograma para el error del conjunto de entrenamiento

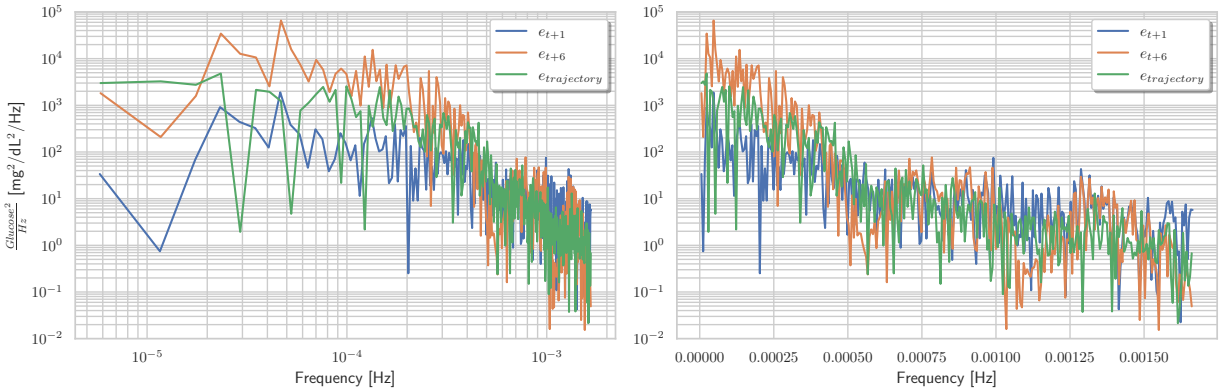

Figura 7: Gráfico del periodograma para el error del conjunto de prueba

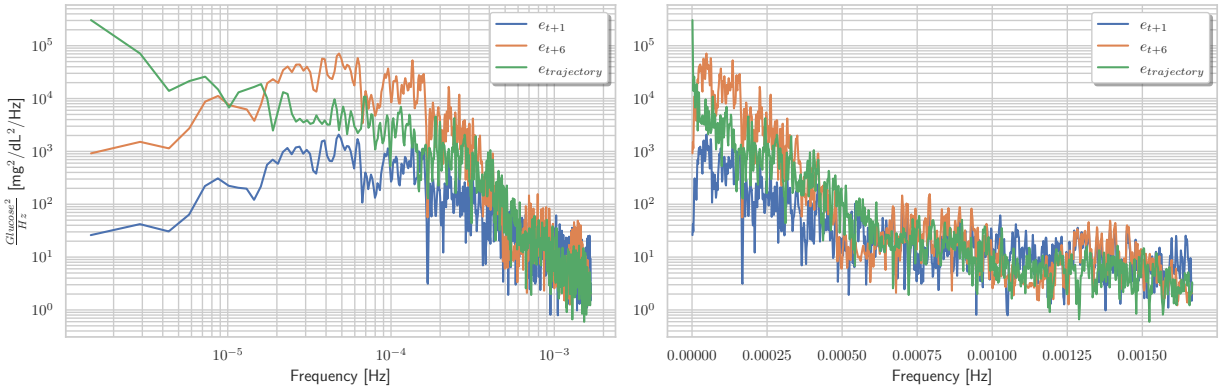

Figura 8: Gráfico de la estimación del espectro para el error del conjunto de entrenamiento

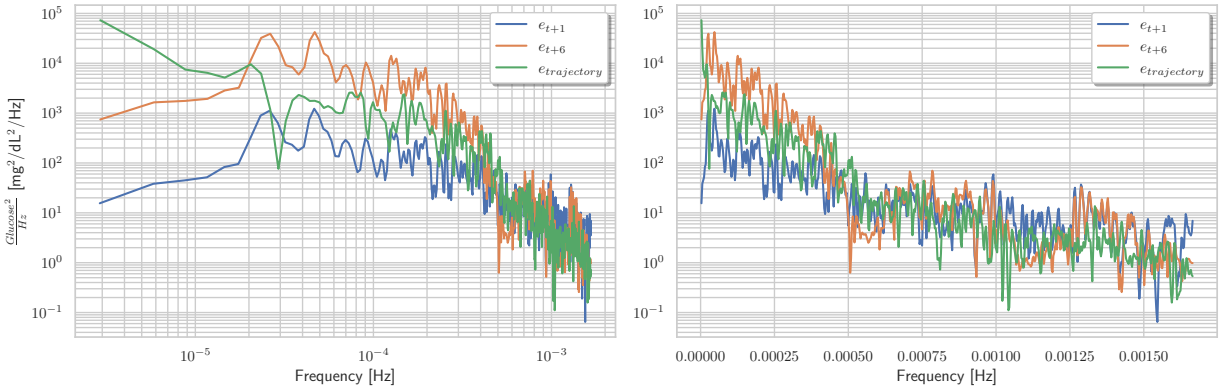

Figura 9: Gráfico de la estimación del espectro para el error del conjunto de prueba

Métricas

Los resultados de las métricas obtenidas bajo este método se muestran en la tabla 3. Hasta el momento, no se ha calculado TG ni ESOD-n, dado que falta estudiar a profundidad su utilidad y si es útil definirlas para la predicción de un paso adelante o la trayectoria total.

	Entrenamiento			Prueba		
	ϵ_{t+1}	ϵ_{t+6}	$\epsilon_{trajectory}$	ϵ_{t+1}	ϵ_{t+6}	$\epsilon_{trajectory}$
RMSE	6.95	34.88	23.53	6.98	33.17	22.66
TG		x			x	
ESOD-n		x			x	

Cuadro 3: Resumen de métricas