

Support Vector Regression para análisis de autorregulación cerebral

Pedro Pablo Silva Antilef¹[0000–1111–2222–3333]

Universidad de Santiago de Chile, Facultad de ingeniería, Departamento de ingeniería informática, Estación Central, Santiago de Chile pedro.silva@usach.cl

Abstract. En este documento se presenta una regresión y análisis de autorregulación cerebral utilizando *Support Vector Regression*. Para esto se generó un modelo de series de tiempo donde la entrada del modelo corresponde a la Presión Arterial Media (PAM) y como salida la Velocidad del Flujo Sanguíneo Cerebral (VFSC). Los datos corresponden a dos pacientes distintos, los cuales pueden estar o no expuestos a una ingesta de CO₂, lo que produce un estado de afección llamado Hipercapnia, lo que debilita la respuesta de la autorregulación cerebral. Es por esto que a los mejores modelos encontrados se les simula un escalón de presión para ver su respuesta (autorregulación cerebral). Finalmente se encontraron modelos con buenos resultados, donde se pudo inferir que ambos pacientes estaban en un estado de afección.

Keywords: Support Vector Regression · Data Mining.

1 Introducción y estado del arte

La autorregulación cerebral es un proceso vascular de alta reactividad, y es el que permite a la circulación cerebral mantener un flujo sanguíneo cerebral estable, a pesar de cambios en la presión arterial de la sangre que llega al cerebro [1]. Se considera también como un mecanismo neuroprotector que en caso de cambios importantes en el tejido cerebral requieren de una compensación inmediata para lograr mantener el flujo de sangre del cerebro [2].

Muchos factores locales pueden afectar la regulación y autorregulación del flujo sanguíneo cerebral. El cerebro demanda insumos esenciales para su funcionamiento como lo son el oxígeno y la glucosa, los cuales deben ser suministrados de manera constante. Esta dependencia continua genera que algunos factores se encuentren presentes de manera simultanea para asegurar este abastecimiento, como lo son factores químicos (metabólicos), miógenos (que da origen a tejido muscular, o que se origina en la musculatura) y neurogénicos (asociado a las neuronas) [2].

La presencia elevada de CO₂ en la sangre arterial se denomina hipercapnia, y se produce generalmente cuando la ventilación alveolar (intercambio de gases entre los alvéolos y el ambiente externo) no logra aumentar adecuadamente como reacción a un aumento de la producción de CO₂, lo que produce una dilatación de las arterias cerebrales con el fin de adecuar la ventilación alveolar,

esto genera un aumento del flujo sanguíneo. En este sentido, la autorregulación cerebral no responde de forma correcta cuando se encuentra en un estado de afección como lo es la hipercapnia, por lo que es posible simular un estado de afección suministrando CO₂ a un paciente con el fin de obtener datos y estudiar la autorregulación cerebral [3].

En el presente texto, se busca analizar la autorregulación cerebral a partir de una regresión que considera como entrada la Presión Arterial Media (PAM) y como salida la Velocidad de Flujo Sanguíneo Cerebral (VFSC), para luego simular un escalón de presión que permita identificar si la autorregulación cerebral responde de manera adecuada. Esta relación ya ha sido estudiada en [4], donde se utilizaron señales PAM y VFSC, suministrado CO₂ a algunos pacientes, con el fin de simular un estado de hipercapnia, para ver cómo se comporta la autorregulación cerebral ante un escalón de presión.

2 Datos

Para esta experiencia se utilizan dos dataset, correspondiente a series de tiempo, con información de pacientes sometidos a mediciones de presión arterial media (PAM) y de velocidad del flujo sanguíneo cerebral (VFSC). Estas mediciones se toman cada 0.2 segundos, es decir con una frecuencia de 5 Hz. Algunos de los individuos sometidos en este estudio, fueron suministrados con dióxido de carbono para simular un estado de hipercapnia y poder analizar su autorregulación cerebral en este estado de afección.

3 Metodología

La metodología corresponde a la regresión y predicción de series de tiempo utilizando *Support Vector Machine* para regresión o *Support Vector Regression*. De una manera similar a la tarea de clasificación, el algoritmo busca encontrar un hiperplano que represente el comportamiento de los datos (hablando de datos lineales, para datos no lineales es más complicado que un hiperplano), luego el algoritmo construye bandas paralelas que cubran la mayor cantidad de datos, lo que se denomina como vectores de soporte, los datos que quedan fuera de este "tubo" generan un margen ε el cual se utiliza para optimizar la inclinación de la curva para obtener el menor error ponderado. Para esto, se divide cada uno de los *dataset* en dos partes iguales (A y B), y se hacen dos experimentos para cada uno de los pacientes; entrenar con el subconjunto A y testear con el subconjunto B y entrenar con el subconjunto B y testear con A. Es importante recalcar que nunca se debe perder el orden de los datos, ya que al hacer esto, se perdería la información que se puede obtener de la cronología intrínseca de un *dataset* temporal (y por ende dinámico). El método tiene ciertos parámetros de calibración que influirán directamente en el resultado, estos parámetros son: (i) costo, el cual corresponde es una penalización que se aplica al violar el margen, (ii) el parámetro gamma define la influencia de búsqueda en entrenamiento, (iii) nu, al igual que cost, es un *trade-off* entre el sobre ajuste y la generalización,

pero específico del método de regresión ν y (iv) los retardos o lags, lo que corresponde a divisiones en las señales con lo que se busca extraer información a un nivel menor.

Para encontrar los mejores parámetros del método, se realiza un proceso de optimización con el método de grilla, el cual genera el modelo para todas las combinaciones posibles de parámetros, con el fin de encontrar la mejor combinación, lo que deriva en un alto costo computacional, dependiendo de los rangos de búsqueda ingresados al proceso.

En términos prácticos, la fórmula de los modelos consta en la utilización de la Presión Arterial Media (PAM) para modelar la Velocidad del Flujo Sanguíneo Cerebral (VFSC), las cuales fueron normalizadas. Como se mencionó anteriormente, también se realizará una validación cruzada balanceada entre las partes A y B de cada señal de paciente. Finalmente y a modo de formalización, se generan los siguientes experimentos:

1. Paciente 1: entrenando con A y testeando con B
2. Paciente 1: entrenando con B y testeando con A
3. Paciente 2: entrenando con A y testeando con B
4. Paciente 2: entrenando con B y testeando con A

Se utiliza la correlación de Pearson como métrica de eficiencia, con lo que se encuentra el mejor modelo de los generados en la grilla de búsqueda de hiperparámetros.

Finalmente, los mejores modelos son testeados en un escalón de presión para analizar su respuesta y poder definir si la serie de tiempo corresponden a un paciente normal o a un paciente con algún tipo de afección.

4 Resultados

En esta sección se muestran los resultados obtenidos de los experimentos detallados en la sección anterior.

Para encontrar los mejores hiperparámetros con el método de grilla se utilizaron distintos rangos para cada uno de ellos, los cuales se pueden observar en la Tabla 1.

Parámetro	Valor
Lag	1, 2, 3, 4, 5
Cost	2^{-4} , 2^{-3} , ..., 2^{12}
ν	0.1, 0.2, ..., 0.9
Gamma	2^{-4} , 2^{-3} , ..., 2^{12}

Table 1: Rangos de búsqueda de parámetros en método de grilla

Utilizando los rangos anteriores se obtuvo que los mejores modelos se encontraron entrenando con la parte A y testeando con la parte B para el caso del

primer paciente y vice versa para el paciente 2. Los mejores valores encontrados con el método de grilla se pueden ver en la Tabla 2.

Experimento	Paciente	Entrenamiento	Testeo	Cost	Nu	Gamma	Correlación
1	1 (G1.001)	A	B	64	0.8	0.0625	0.6365
2	1 (G1.001)	B	A	16	0.8	0.0625	0.6780
3	2 (G1.002)	A	B	0.125	0.5	0.0625	0.5666
4	2 (G1.002)	B	A	4096	0.8	0.25	0.1869

Table 2: Mejores valores encontrados para los parámetros a través del método de grilla

A continuación se muestra cómo el modelo se ajusta a los datos reales de Velocidad de Flujo Sanguíneo Cerebral, tanto para el mejor modelo para el paciente 1, como para el mejor modelo para el paciente 2.

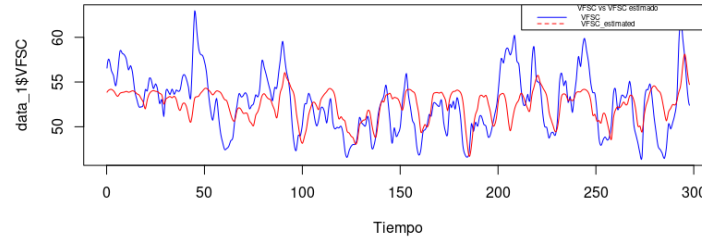


Fig. 1: Gráfico VFSC vs VFSC estimado del mejor modelo para el sujeto 1

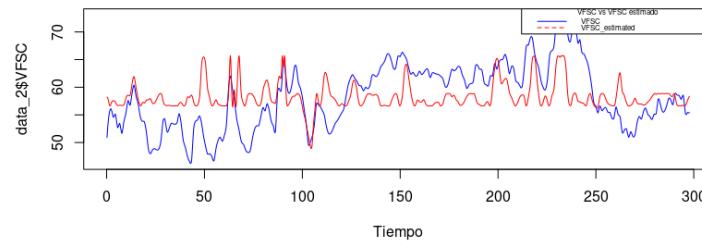


Fig. 2: Gráfico VFSC vs VFSC estimado del mejor modelo para el sujeto 2

Finalmente se muestra la simulación de un escalón de presión, y cómo se comporta el modelo de acuerdo a este cambio repentino, lo que sería un *proxy* de la reacción de autorregulación cerebral.

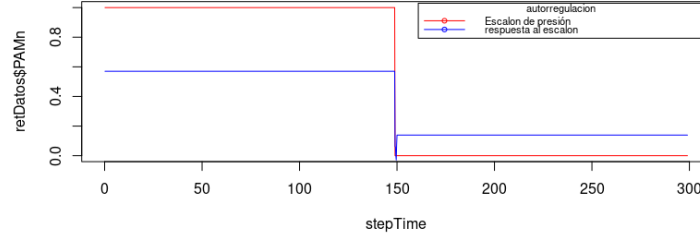


Fig. 3: Respuesta de la VFSC a un escalón de presión para el modelo para el sujeto 1

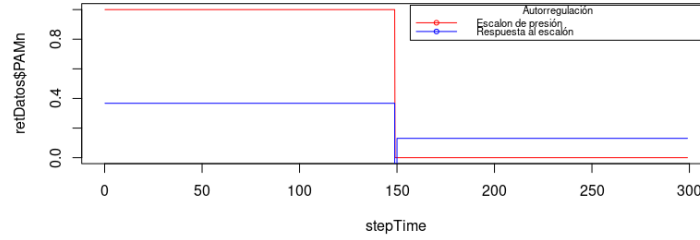


Fig. 4: Respuesta de la VFSC a un escalón de presión para el modelo para el sujeto 2

5 Discusión y conclusiones

El método de grilla logró retornar los modelos con la mejor métricas de eficiencia, para los 4 experimentos, encontrándose que el mejor modelo para el paciente 1 es aquél que se entrena con la parte B y se testea con la parte A. Para el caso del paciente 1, el mejor modelo fue aquel que se entrenó con la parte A y se testeó con la parte B. Es interesante ver que el modelo para el paciente 2 entrenado con B y testeado con A, tuvo un desempeño muy bajo respecto a la correlación de Pearson, probablemente porque la parte B de los datos del paciente 2 no siguen

una tendencia tan clara como los otros segmentos, ya que presenta una subida considerable de la PAM, como se puede ver en la Figura 2.

Si analizamos los rangos de hiperparámetros encontrados, podemos ver que en los experimentos 1, 2 y 3 se obtuvieron valores que se encuentran en el umbral del rango definido para el parámetro Gamma, por lo que se procedió a afinar los rangos de búsqueda para los mejores modelos (experimento 2 y 3), específicamente expandiendo los rangos de gamma en el límite inferior hasta 2^{-6} . Al hacer esto se obtuvieron nuevas combinaciones de valores, los que se muestran en la Tabla 3.

Experimento	Paciente	Entrenamiento	Testeo	Cost	Nu	Gamma	Correlación
2	1 (G1.001)	B	A	32	0.8	0.03125	0.6780
3	2 (G1.002)	A	B	8	0.4	0.01562	0.5776

Table 3: Valores encontrados con el método de la grilla considerando rangos refinados

Estos nuevos modelos no presentaron una mejora para el experimento 2 en términos de la correlación, pero se logró encontrar un valor para Gamma que no se encuentra en los límites de búsqueda de la grilla, por lo que se aceptan los rangos de búsqueda para este experimento. Para el experimento 3 se logró mejorar levemente la correlación, pero el valor de Gamma sigue situado en el límite inferior de los rangos de búsqueda, por lo que se podría seguir ampliando el rango de búsqueda hacia números menores, lo cual no se realizó debido a los importantes tiempos de computación en la búsqueda por grilla. Otro aspecto importante es que los retrasos utilizados no influyeron en la correlación de los modelos, si bien influían en la respuesta al escalón de presión, no dejaba de mostrar el comportamiento de un individuo en un estado de afección, por lo que se optó por elegir los modelos con la mayor cantidad de retrasos, es decir, 5 retrasos.

Si se analiza la Figura 1, podemos ver que en general la VFSC estimada por el mejor modelo para el paciente 1 (Exp. 2) se ajusta de buena manera a los datos crudos, a excepción de algunas variaciones importantes, como los presentes en el tiempo 50, y otros cercanos al tiempo 200 y 250. Algo parecido sucede con el experimento 3 (Figura 2), donde el estimado no logra adaptarse a las variaciones más significativas de la señal. Si analizamos las versiones con rango refinado de estos experimentos podemos observar que para el sujeto 1 (Figura 5), se obtiene un resultado prácticamente igual, a diferencia de lo que sucedió con la versión con rango refinado para el paciente 2, donde la curva estimada se "suavizó", es decir se generalizó el ajuste a los datos, por lo que si bien la correlación es mayor, los datos no se ajustan tan bien a los datos reales.

Si vemos los gráficos de simulación de un escalón de presión con los mejores modelos, podemos ver que ambos parecen tener una respuesta deficiente al cambio de presión (Figura 3 y 4), es decir no lograron una reacción satisfactoria

al cambio de presión, por lo que se podría inferir que ambos pacientes estaban en un estado de afección inducida por la ingesta de CO₂. Estas respuestas no cambiaron cuando se refinó la búsqueda de hiperparámetros (Figura 5 y 6).

6 Anexos

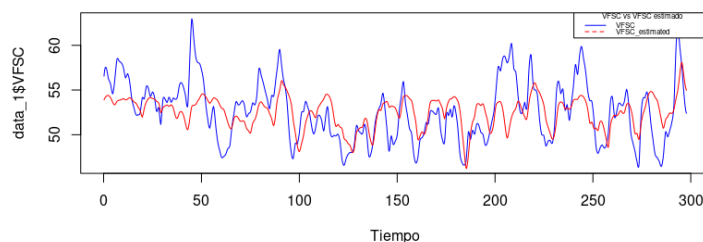


Fig. 5: Gráfico VFSC vs VFSC estimado del mejor modelo con rango refinado para el sujeto 1

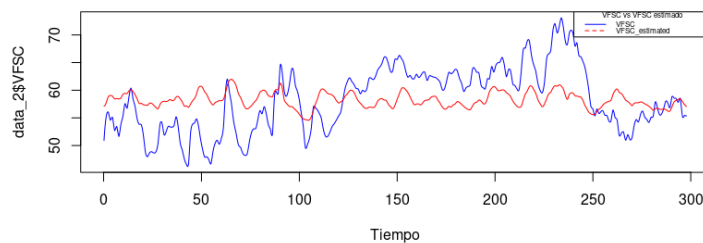


Fig. 6: Gráfico VFSC vs VFSC estimado del mejor modelo con rango refinado para el sujeto 2

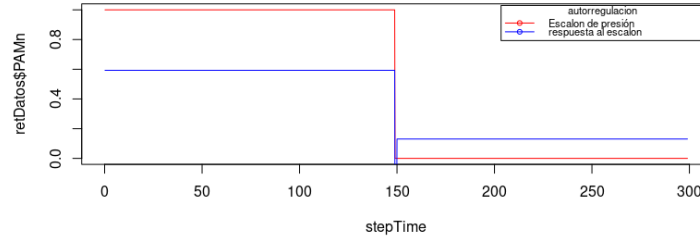


Fig. 7: Respuesta de la VFSC a un escalón de presión para el modelo con rango refinado para el sujeto 1

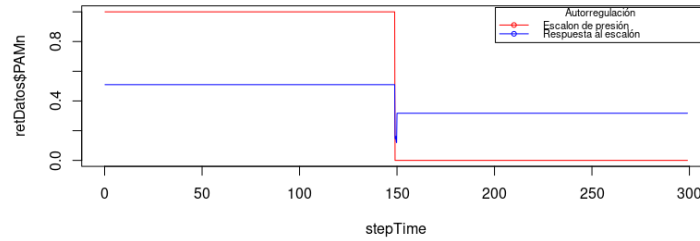


Fig. 8: Respuesta de la VFSC a un escalón de presión para el modelo con rango refinado para el sujeto 2

References

1. Corina Puppò: Introducción Reactividad cerebrovascular. Actualización en el Paciente Crítico. Departamento de Emergencia. Hospital de Clínicas. Uruguay. (2016)
2. Ortiz-Prado, E., Banderas-León, A., Unigarro, L., & Santillan, P. (2018). Oxigenación y flujo sanguíneo cerebral, revisión comprensiva de la literatura. Revista Ecuatoriana de Neurología, 27(1), 80-89.
3. Bhakti K. Patel: Insuficiencia respiratoria hipoxémica aguda (AHRF, ARDS). University of Chicago (2020).
4. Panerai, R. B., Chacon, M., Pereira, R., & Evans, D. H. (2004). Neural network modelling of dynamic cerebral autoregulation: assessment and comparison with established methods. Medical engineering & physics, 26(1), 43-52.