



# EVALUACIÓN DE TÉCNICAS DE APRENDIZAJE POR REFUERZO PARA LA ADAPTACIÓN DE CONTROLADORES DE PÁNCREAS ARTIFICIAL

María Cecilia Serafini<sup>1,2</sup>, Nicolás Rosales<sup>1</sup>, Fabricio Garelli<sup>1</sup>

<sup>1</sup>Grupo de Control Aplicado (GCA), Instituto LEICI (UNLP-CONICET), La Plata, Argentina;

<sup>2</sup>Comisión de Investigaciones Científicas de Buenos Aires, CICpBA, La Plata, Argentina

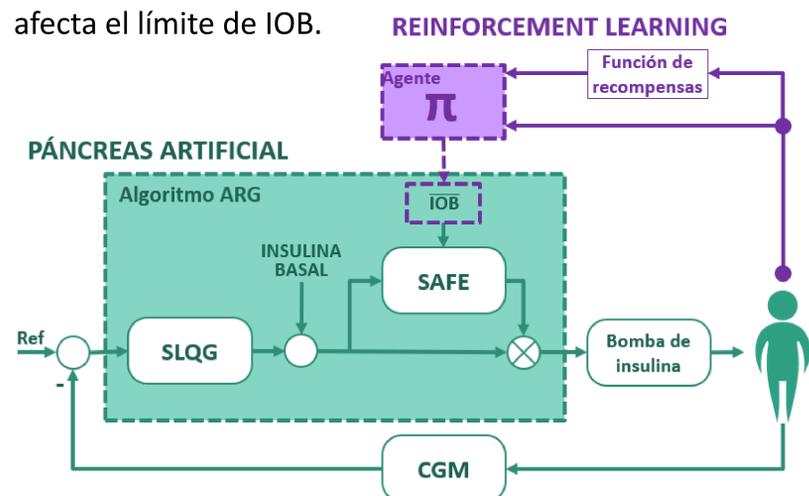


## RESUMEN

El control a lazo cerrado de glucemia en personas con diabetes se ha convertido en un problema emblemático de control automático en la última década. Si bien predominan los sistemas de control híbrido en los que el sistema realimentado se combina con una acción feed-forward para la compensación de ingestas, existen auspiciosos resultados utilizando controladores de lazo cerrado puro. Tal es el caso del controlador ARG que ha sido evaluado en ensayos clínicos en nuestro país. En cualquier caso, el diseño y la evaluación de los controladores se ha centrado en el desempeño de corto plazo, principalmente post-prandial. En este trabajo, se evalúa el potencial de las herramientas de aprendizaje por refuerzo (RL por sus siglas en inglés) para la adaptación de mediano y largo plazo de los controladores glucémicos. En particular, se propone una adaptación basada en Q-learning, una de las técnicas más desarrolladas de aprendizaje por refuerzo sin modelo. La estrategia de adaptación resultante se ilustra en primer lugar sobre un ejemplo sencillo, y luego se aplica a la adaptación del límite de IOB (Insulina a Bordo) del controlador ARG. Los resultados obtenidos, aunque todavía preliminares, permiten concluir que las herramientas de aprendizaje por refuerzo pueden ser de gran utilidad tanto para la adaptación de controladores de lazo cerrado como para su aplicación en sistemas de soporte de decisiones terapéuticas.

## ESTRATEGIA PROPUESTA

- Un agente de RL se utiliza para modificar el límite de IOB a través de una ganancia proporcional K.
- La ganancia se inicializa aleatoriamente en cada episodio.
- El agente fue entrenado usando los 10 pacientes adultos del simulador UVA/Padova.
- Cada episodio equivale a 3 pasos.
- Cada paso es un período de 24hs que incluye 3 comidas de 200 gCHO en total: desayuno a las 7am, almuerzo a las 12 y cena a las 8pm con 60,80 y 60 gCHO respectivamente.
- Observaciones: Tiempo en hipo e hiperglucemia en las 24hs pasadas, discretizado.
- Acciones: Cambios porcentuales sobre la ganancia K que afecta el límite de IOB.

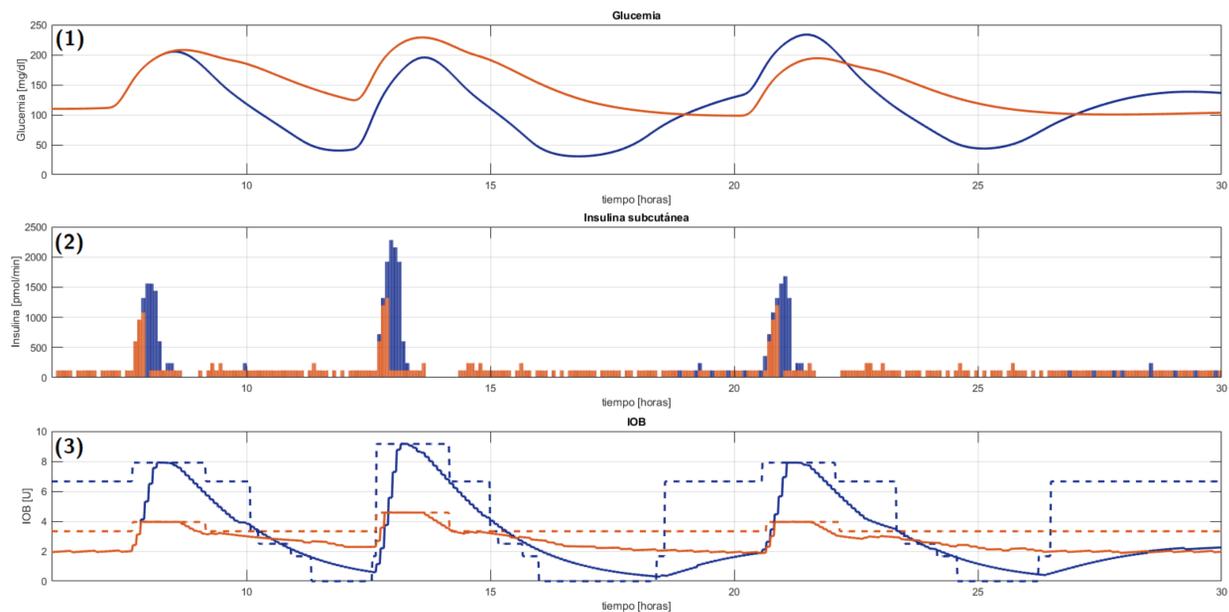


## RESULTADOS

K inicial	Persona	Hipo inicial	Hipo final			Episodios				
			Policy	$\pm 10\%$	$\pm 20\%$	$\pm 50\%$	Policy	$\pm 10\%$	$\pm 20\%$	$\pm 50\%$
1.64	adult#01	20.47	0	0	0	0	2	3	2	3
	adult#02	21.1	0	0	0	0	2	4	3	2
	adult#03	13.67	0	0	0	0	2	3	2	3
	adult#04	25	0	0	0	0	2	5	3	8
	adult#05	25	0	0	0	25	3	4	17	>20
	adult#06	18.6	0	0	18.6	18.95	2	>20	>20	>20
	adult#07	25	0	0	19.78	25	4	>20	>20	>20
	adult#08	25	0	5.69	0	0	2	>20	>20	>20
	adult#09	25	0	0	19.57	25	2	>20	>20	>20
	adult#10	22.97	0	1.39	0	0	2	4	4	2

**Tabla:** Comparación entre resultados de sintonización del límite de IOB vía RL y un esquema manual. Tiempos en hiper e hipo glucemia antes y después de la sintonización y número de episodios necesarios para alcanzarla.

El agente de RL logra eliminar las hipoglucemias por completo para todos los pacientes, mientras que los esquemas manuales ni siquiera logran alcanzar un valor estable de la ganancia K, en muchos casos.



**Figura:** Evolución de la excursión de glucosa (1), infusión de insulina (2) y IOB (3) para el Adulto#04 del simulador UVA/Padova, utilizando el algoritmo ARG con (naranja) y sin (violeta) sintonización RL. Abajo: IOB (línea sólida) y límite de IOB (línea punteada)

Se puede ver que la infusión de insulina posprandial se limita de una forma que evita los episodios hipoglucémicos a la vez que reduce la excursión glucémica, sin incrementar significativamente los valores máximos.

## CONCLUSIONES Y TRABAJO FUTURO

En este trabajo se analizaron los alcances y limitaciones de la técnica de Q-Learning para su aplicación en la adaptación de controladores de un sistema de páncreas artificial. Los resultados muestran que es una herramienta con gran potencial, ya que el control glucémico se trata de un problema que requiere limitaciones fuertes sobre determinadas condiciones (como por ejemplo, evitar hipoglucemias severas).

Como trabajo futuro se podría considerar la adaptación de otros parámetros del controlador (ej. Ganancias agresiva y conservativa), su utilización en sistemas de soporte de decisión para pacientes con terapia MDI o con bomba de insulina a lazo abierto, entre otras aplicaciones. Asimismo, se espera poder explorar el efecto de la definición de recompensas para un balance más equilibrado entre hipo e hiperglucemia

sin comprometer la excursión glucémica y considerar esquemas de discretización incremental o estados/acciones continuas de manera de poder graduar la agresividad y los pasos de la adaptación

## Referencias:

- [1] P. Colmegna, F. Garelli, H. D. Battista y R. Sánchez-Peña, «Automatic regulatory control in type 1 diabetes without carbohydrate counting,» Control Eng Pract, nº 74, pp. 22-32, 2018.
- [2] R. S. Sutton and A. G. Barto, Reinforcement Learning: An Introduction, 1st ed. MIT Press, 1998.
- [3] M. De Paula, L. O. Ávila et al., «Controlling blood glucose variability under uncertainty using reinforcement learning and Gaussian processes,» Appl. Soft Comput. J., vol. 35, p. 310, 2015.