Buffer System Control for Hybrid Plants

M. De Paula, C.R. Sanchez Reinoso, member IEEE y L.O. Avila

***Abstract*— The use of multimodal control techniques for batch dynamic systems offers a promising alternative to successfully perform tasks of supervision and control in industrial processes. These techniques seek to integrate different control strategies with partial objectives or "behaviors" using Lebesgue automatons overlooking achieve certain operational objectives. These automatons can identify optimal sequences of modes, also called control programs, using system simulations. Multimodal control programs consist of a sequence of modes, each of which comprises a feedback control law $\kappa(x)$ and relevant termination conditions $\xi(x,T)$. In this paper the use of a reinforcement learning algorithm is described via to find, by simulations, optimal control program aligned to maximize productivity in a hybrid process plant composed oftwo buffer tanks.**

***Keywords*— batch systems, reinforcement learning, multimodal control, Lebesgue sampling.**

I. INTRODUCCIÓN

En muchos de los sistemas de producción en las plantas químicas existen sistemas con dinámicas híbridas [1], [2]. Estos aparecen en los sectores donde es necesario acoplar operaciones por lotes, como reactores o cristalizadores, con otra continua como los trenes de separación u operaciones de secado [3]. La función de integración entre estas distintas formas de operación se realiza por medios inventarios intermedios o tanques buffer. Una problemática importante radica en la utilización de la capacidad de estos tanques de tal suerte que no se limite innecesariamente la productividad del proceso en conjunto.

El control óptimo de los sistemas dinámicos híbridos es un problema que ha recibido mucha atención por los investigadores. Principalmente, la optimización y el control de los sistemas híbridos se han abordado mediante el desarrollo de modelos que describen dinámicas que combinan estados continuos y discretos [4]–[6]. De esta manera, el comportamiento de un sistema dinámico híbrido en tiempo continuo puede ser representado como una secuencia de evoluciones continuas intercaladas por eventos discretos. A cada una de estas evoluciones continuas se las puede denominar como “comportamiento” o “modo de operación” del sistema [7], [8].

Para un número finito de modos de operación se planteó un novedoso paradigma de modelado conocido como autómata híbrido en tiempo continuo, denominado en inglés como integral continuous-time hybrid automata (icHA) [9]. Este paradigma propone un control de eventos en base a la formulación y resolución de un problema de optimización en el que no se asume información a priori acerca del momento y orden de los eventos. Desde la perspectiva del control óptimo no lineal, cuando la programación dinámica es aplicada a sistemas híbridos en tiempo continuo (integral continuous-time hybrid systems), el problema computacional resultante es altamente costoso dado que usualmente abarca programas matemáticos no convexos [10]. Muchos trabajos previos relacionados con la aplicación del formalismo de programación dinámica al control óptimo y planificación de secuencia de modos para sistemas con lógicas de conmutación (switched systems), revelan la cuestión de los problemas de optimización combinatoria con complejidad exponencial [11]–[13].

La solución de los problemas de optimización dinámica para sistemas híbridos en tiempo continuo [14] ha sido profusamente revisada por Barton y Colab. [2] donde los autores proponen un enfoque para dividir al problema original en otros problemas más simples, como por ejemplo determinar la secuencia óptima de modos de operación asumiendo tiempos de transición fijos entre estados discretos.

Desde la perspectiva del control basado en el comportamiento, tomado del campo de la robótica, Mehta y Edgerstedt [15] también trataron el importante problema de identificar secuencias óptimas de modos en el control óptimo multimodal usando técnicas de aprendizaje por refuerzos [16]–[18]. El aprendizaje por refuerzos [16] es una técnica conveniente para resolver problemas de control óptimo bajo incertidumbre y demuestra ser efectiva para abordar el problema de concatenar o combinar modos de control con leyes de retroalimentación y condiciones de terminación pre-establecidas. Para sobrellevar el inconveniente de un número fijo de modos, Mehta y Egerstedt proponen aumentar el conjunto de modos como una función de los modos actuales usando cálculo variacional [19].

Un sistema dinámico híbrido modelado como sistema dinámico multimodal es un sistema controlado cuya dinámica alterna entre un conjunto finito de posibilidades con vista a desplegar un dado comportamiento o alcanzar un objetivo de interés. Estos cambios del comportamiento pueden ser el resultado de un evento específico o una decisión planificada. La estrategia de control se resume en implementar un programa óptimo de modos que permita alcanzar un determinado objetivo derivado del estado deseado para un sistema o proceso a pesar de las perturbaciones que alteran el curso y resultado de cada modo aplicado. Uno cualquiera de estos programas de control (π) consiste en una dada secuencia finita de modos (σ_i), donde cada uno de estos modos está compuesto de una ley decontrol $\kappa^i(x)$ y sus condiciones relevantes de terminación $\xi^i(x,T)$ [19], [20].

M. De Paula, Universidad Nacional del Centro, Buenos Aires, Argentina, mariano.depaula@fio.unicen.edu.ar

C. R. Sanchez Reinoso, Consejo Nacional de Investigaciones Científicas y Técnicas, Catamarca, Argentina, csanchezreinoso@santafe-conicet.gov.ar

L. O. Avila, Laboratorio de Investigación y Desarrollo en Inteligencia Computacional, CONICET-UNSL, San Luis, Argentina, loavila@unsl.edu.ar

Corresponding author: Carlos Sanchez Reinoso

En este trabajo se presenta un algoritmo de aprendizaje por refuerzos que permite identificar la secuencia óptima de modos usando simulaciones del sistema estudiado. Se aborda un caso de estudio relacionado con la interfaz entre un conjunto de reactores discontinuos que abastecen un tren de separación del producto. El sistema bajo estudio está compuesto por dos tanques, mientras que el objetivo de control es maximizar la productividad del proceso químico híbrido en cada campaña de producción. Las leyes de control, de cada modo, son diseñadas para alcanzar una determinada meta parcial de acumulación o drenaje del inventario intermedio.

I. Programas de control multimodal

Supóngase que la dinámica del estado *x* responde a:

$$\frac{dx}{dt} = f(x,u(t),z(t)), x \in X = R^n, u \in U = R^m \quad (1)$$

de donde *z(t)* es una perturbación medible que evoluciona en el tiempo según:

$$\frac{dz}{dt} = g(z,t), z \in Z = R^d \quad (2)$$

Si en un determinado momento τ_0, en el que el estado del sistema es $x(\tau_0)$ y la perturbación es $z(\tau_0)$, el sistema recibe una secuencia de modos $\pi = \{(\kappa_1, \xi_1),\ldots,(\kappa_q, \xi_q)\}$ se desencadena la transición de estados:

$$\delta(\sigma, x(\tau_0), z(\tau_0)) = x(\tau_0) + \int_{\tau_0}^{\tau_1} f(x(t),u(t),z(t)dt + \ldots$$

$$\ldots + \int_{\tau_{q-1}}^{\tau_q} f(x(t),u(t),z(t)dt \quad (3)$$

Si la secuencia de modos π posee una longitud acotada, sólo será posible alcanzar un conjunto finito de estados. Por tanto la aplicación de programas de control con una longitud máxima *N*, resulta en una cuantización de los estados del sistema. De acuerdo con esto se obtiene una discretización finita del espacio de estados, conocido como *"Lebesgue-sampled state machine"* $(X_N^Q, \Sigma, \tilde{\delta}, \tilde{x}_0, \tilde{z}_0)$[19]–[21]. El superíndice *Q* significa que estos estados pertenecen al espacio de función valor o utilidad[16], mientras que la función de transición de estados $\tilde{\delta}$ es:

$$\tilde{x}_{k+1} = \tilde{\delta}(\tilde{x}_k, \tilde{z}_k, \sigma_k) = \delta(\sigma, x(\tau_0), z(\tau_0)), k = 0,1,2,\ldots,$$

$$x_0 = x(\tau_0), z_0 = z(\tau_0), \quad (4)$$

El espacio de estados discreto X_N^Q está dado por el conjunto de todos los estados que pueden alcanzarse desde $\chi_0 = (x_0, z_0)$ y para una dada evolución del vector de perturbaciones *z(t)* cuando se implementa una secuencia de modos $\sigma \in \sum$ de longitud menor o igual que *N*. En el marco del control óptimo, el objetivo consiste en encontrar una secuencia de modos tal que maximice la acumulación de *rewards* obtenidos en las transiciones de estado del autómata híbrido de Lebesgue, el cual se esboza su idea en la Fig. 1. De acuerdo con la discretización multimodal X_N^Q del espacio de estados y la dinámica de transición del autómata híbrido de Lebesgue es posible emplear el paradigma de aprendizaje por refuerzos (*reinforcement learning*) para estimar los *Q-values* correspondientes a la función de *premios (castigos)* relacionada con el objetivo de control[16]. Durante el proceso de aprendizaje, es ventajoso que los estados explorados y los escenarios de perturbación sean utilizados para aprender, o descubrir, cuál es el mejor modo de operación σ^* para cada par $\chi \in X_N^Q$.

En el pseudocódigo del Algoritmo 1 la discretización inicial del espacio estado-perturbación X_N^Q se asume desconocida. El proceso de aprendizaje consiste en simular las posibles transiciones hacia todos los estados alcanzables, partiendo desde un dado estado-perturbación $\tilde{\chi}_0 = (\tilde{x}_0, \tilde{z}_0)$ y usando los distintos modos que son factibles de aplicar en el estado inicial. En cada iteración del proceso de aprendizaje un par estado-perturbación se elige aleatoriamente del conjunto de estados-perturbaciones $\aleph$ visitados. Luego se aplica uno de los modos de control del conjunto de modos posibles Σ, lo cual genera la transición al siguiente estado $\tilde{x}'$. En dicha transición se obtiene un premio o *reward* $r(\tilde{\chi}', \sigma_i)$ que se consigue una vez finalizada la ejecución del modo σ_i, momento en el cual ξ_i=1.

En el Algoritmo 1, la función *step*($\tilde{x}$) representa la longitud del programa de control más corto necesario para alcanzar un estado $\tilde{x}'$ a partir de un estado inicial $\tilde{x}_0$. De esta forma, son explorados únicamente los estados resultantes de aplicar una secuencia de modos de longitud menor o igual que *N*, esto garantiza que $\tilde{x}' \subset \tilde{\chi}' \in X_N^Q$. En cada transición a un nuevo estado $\tilde{x}'$, es necesario determinar si este estado pertenece o no al conjunto de estados visitados previamente. En el caso de que $\tilde{\chi}'$ no pertenezca al espacio de estados debe incorporarse al mismo, incrementando *step*($\tilde{x}$) en 1, y una nueva entrada debe ser agregada en la tabla *Q*. En las situaciones que $\tilde{\chi}'$ pertenece al entorno de un estado visitado previamente, se actualiza el valor *Q* del estado que define el entorno en cuestión. La exploración de estado-perturbación continúa de esta forma hasta que se estabilicen las entradas de la tabla *Q* y no aparecen nuevos estados.

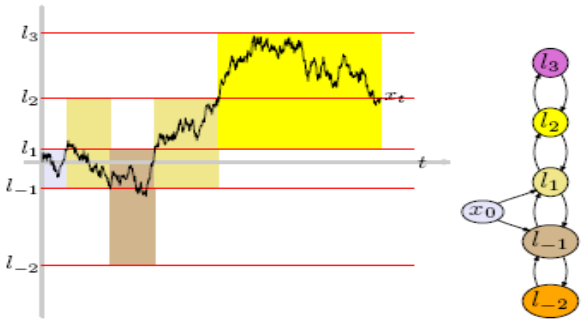


Fig. 1. Autómata híbrido con muestreo de Lebesgue.

1.- $\aleph := \{\tilde{\chi}_0, \delta(\tilde{\chi}_0, \sigma)\}$; $step(\tilde{\chi}_0) := 0$;
2.-$step\ \tilde{\delta}(\tilde{\chi}_0, \sigma) := 1$; $\forall \sigma \in \Sigma$
3.-k:=1; *índice para el recuento de visitas a pares estado de perturbación* $\chi \in \aleph$
4.-$Q_k(\tilde{\chi}, \sigma) := const, \forall \chi \in \aleph, \sigma \in \Sigma$
repeat
5.- k=k+1
6.- $(\tilde{\chi}) := rand(\chi \in \aleph | step(\chi) < N)$
7.- $\sigma := rand(\Sigma)$
8.- $\tilde{\chi}' := \tilde{\delta}(\tilde{\chi}, \sigma)$
9.-**if** $\tilde{\chi}' \notin \aleph$ **then**
10.- $step(\tilde{\chi}') = step(\tilde{\chi}) + 1$
11.- $\aleph := \aleph \cup \tilde{\chi}'$
12.- $Q_k(\tilde{\chi}', \sigma) := const, \forall \chi \in \aleph, \sigma \in \Sigma$
end if
12.- $Q_{k+1}(\tilde{\chi}', \sigma) := Q_k(\tilde{\chi}', \sigma) + \alpha_k [\rho(\tilde{\chi}', \sigma) +$
$+ \gamma \max_{\sigma' \in \Sigma} [Q_{k-1}(\tilde{\chi}', \sigma') - Q_{k-1}(\tilde{\chi}, \sigma)]$
13.- **until** $\text{mod}(k, L) = 0$ and
$|Q_k(\tilde{\chi}, \sigma) - Q_{k-L}(\tilde{\chi}, \sigma)| < \varepsilon, \forall \chi \in \aleph, \sigma \in \Sigma$
$X_N^Q = \aleph$

Algoritmo 1. Pseudocódigo del algoritmo *Q-learning* de control multimodal.

Las condiciones de stop del bucle principal (línea 13) se imponen de tal suerte que las combinaciones de estados-perturbaciones $\aleph$ sean visitados una suficiente cantidad de veces, y además que todos los modos $\sigma \in \Sigma$ sean ensayados repetidas veces para lograr una adecuada convergencia de las entradas en la tabla Q.

II. Caso de Estudio

A. Descripción general

En diversos procesos de la industria química tales como manufactura de PVC o la industria azucarera existen dispositivos de almacenamiento temporario (*buffers*) que se emplean con el fin de acoplar, sin pérdida de productividad, operaciones discontinuas como reactores o cristalizadores con operaciones de naturaleza continua como secaderos, trenes de destilación, etc.[22], [23].

En el problema a resolver consideremos un conjunto de n reactores descargando en paralelo a un tanque *buffer* cuya salida alimenta el proceso aguas abajo, como se indica en la Fig.(2). La gestión adecuada de la capacidad de estos tanques no es un problema de fácil solución, sujeto a las restricciones operativas de cada proceso en particular y de la variabilidad del *scheduling* de descarga desde los reactores. Siendo que el objetivo que se persigue es el de maximizar la productividad del proceso en conjunto, es razonable pensar que una buena gestión de la capacidad del inventario intermedio en el tanque buffer consiste en brindar un flujo aguas abajo tal que el caudal volumétrico enviado hacia la sección continua sea el máximo posible en todo momento, bajo una única restricción que es evitar el rebalse o secado el tanque buffer[22].

Pensando el problema desde el punto de vista del control clásico la variable manipulada sería, por ejemplo, el flujo de salida del tanque buffer ($F_{out}(t)$). En este sentido, resulta necesario definir una variable de referencia o *set point*, para lo cual se podría tomar un determinado nivel del tanque como una referencia a seguir, por caso la altura media del tanque. Operado de esta manera, es muy posible que la variable manipulada ($F_{out}(t)$) deba experimentar cambios abruptos teniendo en cuenta la discontinuidad de la descarga de los reactores. A efectos prácticos, esta situación no es deseable debido a que atenta con la productividad y continuidad operativa del conjunto.

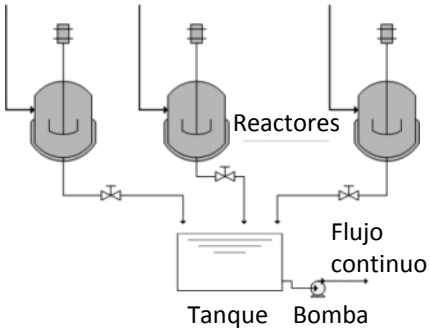


Fig. 2. Interfaz entre las secciones discontinuas y continuas de una planta híbrida.

En la situación expuesta en este trabajo no se fija una referencia o *set point* de antemano para ninguna de las variables (como por ejemplo cantidad de líquido almacenado o flujo de salida del *buffer*), sino que solamente se busca maximizar la productividad del sistema completo evitando eventos inadmisibles como que el tanque buffer rebalse o se seque. De esta manera se busca maximizar la tasa de descarga promedio al mismo tiempo que se busca evitar variaciones abruptas del flujo de descarga instantáneo $F_{out}(t)$ del buffer como respuesta a los cambios de patrones de descarga de los reactores.

En nuestro caso, la variable manipulada es el flujo de salida del tanque buffer ($F_{out}(t)$), cuya magnitud y variación se controla mediante leyes de retroalimentación que deben estar

en función de las variables de estado disponibles en el sistema para cada modo de control.

Hasta aquí, con el problema así definido, no se identifican en el sistema variables de estado para los flujos de entrada, que cumplan con la condición de Markov[6], de tal forma que sea posible aplicar las técnicas de *Q*-Learning antes descriptas.

Con el fin de conseguir variables de estado Markovianas[16], [24] para el flujo total de entrada al buffer se propone una modificación de las condiciones de diseño del sistema buffer (véase Fig. 3). Esto es, en lugar de utilizar un solo tanque buffer, es posible usar dos tanques en serie dispuestos de forma que los *n* reactores descarguen directamente al primero de ellos (de aquí en adelante se llamará como "*tanque suavizador*") y este descargue directamente al segundo ("*tanque controlado*"), según la ley de Torricelli. De esta manera, se tiene un sistema abierto en el que el que el tanque superior se comporta como un filtro de baja frecuencia generando una salida continua que alimenta al tanque *buffer* inferior.

Con esta configuración se consiguen dos variables de estado genuinas que resumen la información del estado del sistema en todo momento. La primera de ellas es la altura del tanque suavizador ($x_1(t)$) y la segunda es la altura del tanque controlado ($x_2(t)$). Las variables $x_1(t)$ y $x_2(t)$ son variables de estados markovianas, dado que en su valor presente reflejan toda la historia previa reciente. Por otra parte, el estado del tanque suavizador depende sólo del patrón de descarga de los reactores, con lo cual constituyen una adecuada representación del cronograma de producción (*schedule*). Lógicamente, sí el patrón de variación de los flujos cambia entonces el patrón de variación de $x_1(t)$ también va a cambiar y, consecuentemente, deberá cambiar el control de $x_2(t)$.

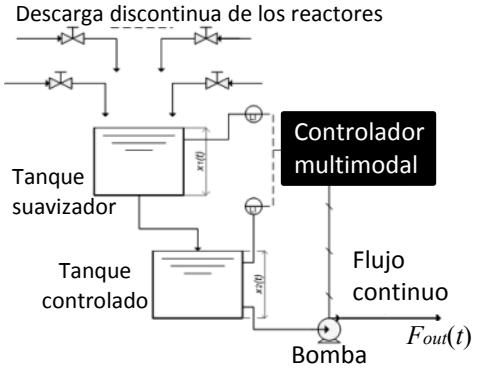


Fig. 3. Sistema con dos tanques.

En los procesos de producción por lotes (procesos *batch*) los *n* reactores trabajan respetando un determinado programa de producción (*schedule*), con lo cual el patrón de descarga del conjunto de reactores está directamente vinculado con la estructura de dicho programa de producción. Así, el flujo de salida del tanque suavizador está directamente relacionado con el patrón de variación del cronograma de las descargas. Este flujo pasa directamente al tanque controlado, cuya salida ($F_{out}(t)$) será manipulada en función de las variables de estado del sistema en pos de conseguir los objetivos propuestos.

La idea de aplicar las técnicas de control multimodal para este caso de estudio consiste en encontrar una secuencia de modos (π) que permitan cumplir los objetivos mencionados para toda la campaña de producción. Como se explicó anteriormente, los modos se definen mediante una serie de leyes de control (σ_i) y sus respectivas condiciones de terminación (ξ_i). Estas leyes y condiciones de terminación actúan directamente sobre las variables de estado del sistema para manipular la variable de control $F_{out}(t)$.

B. Caso Particular

Como caso particular para ilustrar cuantitativamente el enfoque propuesto se considera un sistema constituido por cuatro reactores descargando paralelamente a un tanque suavizador y éste al tanque controlado.

Algunas consideraciones a tener en cuenta, antes de aplicar la técnica para obtener el programa de control, son las relacionadas con aspectos técnicos de diseños de los equipos. El tanque al que descargan los reactores debe estar diseñado para drenar según la ley de Torricelli, en donde el flujo de descarga (*f*) es:

$$f = \gamma\sqrt{h} \tag{3}$$

además dicho tanque deberá tener un volumen suficiente para soportar el patrón de descarga de los reactores, de manera de que no se produzcan desbordes en el suavizador. Para simplificar el problema se asume que los todos los reactores durante su correspondiente fase de descarga, drenan el mismo caudal de forma intermitente según un dado cronograma de descarga (producción). En las siguientes sub-secciones se detallarán las características del sistema de dos tanques y seguidamente se proporcionaran detalles de la implementación de la técnica de control multimodal para el cronograma de variación de los flujos desde los reactores.

C. Especificaciones

El tanque suavizador es un recipiente cilíndrico de 1.5 m^3 de volumen y con 1.5 m de altura útil, que descarga según la Eq. (3) con γ=1; en tanto el tanque controlado posee un volumen de 1 m^3 con una altura máxima de 1.0 m.

Las leyes de control para los distintos modos se definen con las metas parciales de aumentar (acumular) o disminuir (drenar) la altura promedio del tanque controlado. Las condiciones de parada comunes a todos los modos consisten en detenerlos cuando se exceda el nivel máximo del tanque controlado (H_2) o cuando se llegue a un nivel cercano a cero, es decir cuando se rebalsa o se seca el buffer. Para el caso de los modos cuya meta parcial sea lograr una tendencia creciente de la evolución de la altura promedio, la condición de parada del modo se corresponde en detener cuando la imposibilidad práctica de incrementar o disminuir la altura promedio indica la futilidad de continuar la ejecución. A modo de ejemplo, para los modos cuyo objetivo es lograr un decrecimiento en la tendencia de evolución de la altura del tanque, los mismos se detendrán cuando la tendencia de dicha evolución sea creciente. Para lograr un comportamiento suave

del flujo de descarga $F_{out}(t)$, las leyes de control actúan sobre un valor suavizado de la altura instantánea del tanque controlado, de acuerdo al conocido criterio de ajuste exponencial, con un coeficiente de ajuste α=0.01.

D. Aplicación del algoritmo

En primer lugar se define un estado inicial x_0 a un tiempo t_0, en el cual se empiezan a aplicar las acciones de control. Dicho estado se puede establecer arbitrariamente como el par $(x_1(t_0), x_2(t_0))$, para el cual $x_2(t_0)$ alcanza un cierto porcentaje de la altura máxima del tanque inferior, en este ejemplo se establece en un 20% de H_2. Tanto para los modos cuya meta es lograr una tendencia positiva en la evolución de la altura como para los que tienen como meta una evolución con tendencia decreciente, las leyes de control siguen la forma indicada en la Fig. 4.

Los modos 1 y 3 se comportan según la ley de la recta sólida (azul), en tanto que los modos 2 y 4 lo hacen según la recta de trazos (roja). En todos los casos los modos detienen su ejecución cuando se seca el tanque controlado o cuando el nivel de líquido excede su altura máxima posible (H_2), esto es $\xi_i=0 \Rightarrow \xi_i=1$ sí $h_2(t) \leq \varepsilon$ ó $h_2(t) \geq H_2$. Por otra parte, los modos 1 y 2 tienen el objetivo de generar un incremento en la altura del tanque controlado. Para simplificar el ejemplo, dicha tendencia se vincula directamente con la pendiente de la recta de ajuste lineal de los últimos cinco valores de altura suavizada, interrumpiendo la ejecución del modo cuando la misma es negativa. Análogamente, los modos 2 y 4 se detienen cuando el valor de la pendiente de la recta de ajuste sea positivo.

Las leyes de retroalimentación han sido diseñadas de tal forma de responder a las exigencias operativas del proceso al mismo tiempo que buscan variar suavemente el flujo de salida del tanque inferior.

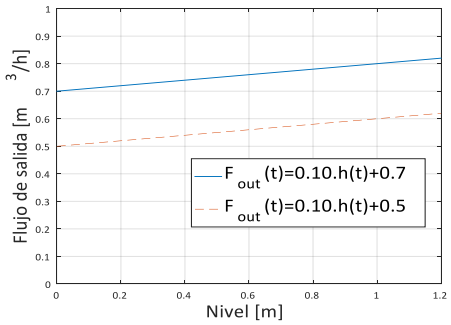


Fig. 4. Leyes de retroalimentación.

En línea con el objetivo planteado, para la aplicación del algoritmo propuesto (Algoritmo 1) la función de *rewards* se define como el volumen descargado durante la aplicación del modo siempre y cuando no se seque ni rebalse el tanque. En este último caso al *reward* se le asigna un valor negativo de -5. De esta forma resulta:

$$r = \int_t^{t+\tau} F_{out}(t)dt \Leftrightarrow x_2(t) < H_2 \wedge x_2(t) > 0$$
$$r=-5 \qquad \Leftrightarrow x_2(t) \geq H_2 \vee x_2(t) \leq \varepsilon \qquad (6)$$

Otra cuestión a definir respecto al punto 9 del Algoritmo 1 es el criterio de "similitud" entre los estados visitados para determinar si un estado pertenece o no al espacio de estados explorado. En este caso se asume que un estado s_t pertenece al espacio de estados previamente visitados (ℵ) si se cumple que:

$$s_t \in \aleph \Leftrightarrow \exists s_i \in \aleph / \| s_t - s_i \| \leq \rho.\|H_2\| \qquad (7)$$

En el caso examinado se asumió ρ = 0.05.

III. Resultados

En el caso de estudio analizado se tiene un programa de producción intermitente para diez lotes de producción consecutivos de cuatros reactores trabajando en paralelo. Para el análisis se toman los resultados de controlar el sistema durante un horizonte temporal de tres mil minutos. En la Fig. 5 se indica el patrón del flujo de descarga del tanque suavizador ocasionado por un determinado programa de producción. La Fig.6 indica la evolución del nivel de líquido en el tanque controlado, como consecuencia de operar el sistema mediante el programa de control π encontrado mediante la aplicación del Algoritmo 1. En la Fig.7 se muestra la evolución temporal del flujo de descarga aguas abajo del tanque controlado la cual está directamente vinculada con la secuencia de modos seleccionada por el programa de control (Fig.8) aprendido.

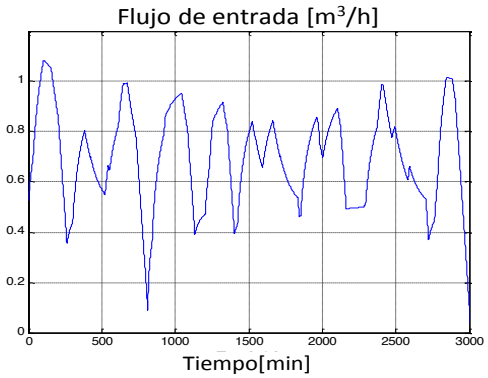


Fig. 5. Flujo de carga al tanque controlado.

Es para destacar que, según se observa en Fig.6, el sistema trata en todo momento de elegir aquellos modos de control que generan una mayor productividad aguas abajo para el proceso.

Como puede observarse en los resultados presentados, la cuestión más importante a resaltar de la aplicación del control multimodal surge al comparar la variación del flujo de entrada al tanque controlado (Fig.5) con la variación en la descarga del mismo (Fig.7). Se puede apreciar que uno de los objetivos planteados se cumple satisfactoriamente: la salida promedio se mantiene elevada a la par que se evitan que los cambios bruscos a la entrada del tanque controlado se trasladen directamente a la salida del mismo. Si bien se observan variaciones en el caudal de descarga, ésta se comporta de manera suave durante períodos prolongados de tiempo y la frecuencia y magnitud de los cambios de caudal de descarga es baja.

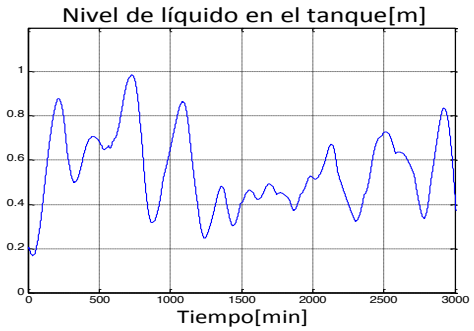


Fig. 6. Evolución de la altura en el tanque controlado.

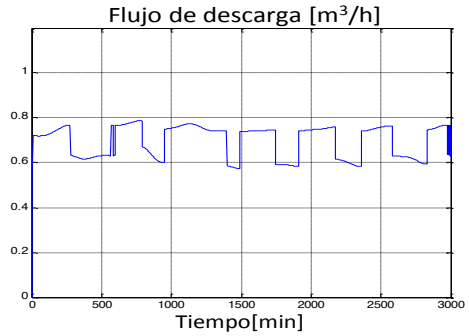


Fig. 7. Flujo de descarga aguas abajo.

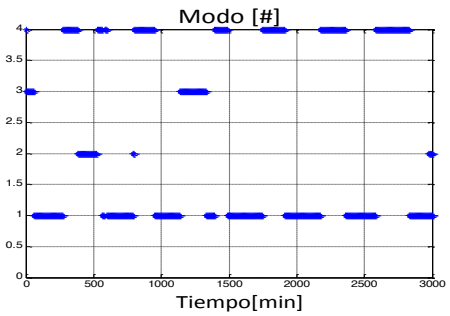


Fig. 8. Secuencia óptima de modos.

IV. Conclusiones

Los resultados obtenidos en simulación demuestran el potencial del enfoque propuesto para abordar los problemas de optimización de la operación de sistemas dinámicos híbridos. Este tipo de comportamiento muy a menudo surge como una consecuencia de operar los sistemas dinámicos continuos con diferentes modos de operación. Para abordar el problema de operación óptima de los sistemas híbridos fue elaborada una propuestas de control multimodal a través del diseño de un novedoso algoritmo de aprendizaje por refuerzo que permite encontrar, al menos, una política óptima de conmutación de modos de control de tal forma de maximizar el rendimiento cumpliendo con las especificaciones (o restricciones) operativas impuestas.

Por otra parte, el hecho de usar una política de control para operar los sistemas dinámicos híbridos representa una ventaja distintiva dado que una política de control puede actuar de manera reactiva y sin necesidad alguna de tener que resolver on-line un programa de optimización para determinar una acción de control. Otra ventaja importante del enfoque propuesto es que permite que la política de control pueda ser continuamente actualizada (o adaptada) en base al comportamiento observado con el fin de tener en cuenta siempre los cambios intrínsecos del comportamiento del sistema en cuestión.

Referencias

[1] S. Engell, S. Kowalewski, C. Schulz, and O. Stursberg, "Continuous-discrete interactions in chemical processing plants," *Proc. IEEE*, vol. 88, no. 7, pp. 1050–1068, Jul. 2000.

[2] P. I. Barton, C. K. Lee, and M. Yunt, "Optimization of hybrid systems," *Comput. Chem. Eng.*, vol. 30, no. 10–12, pp. 1576–1589, Sep. 2006.

[3] R. Peirce and S. Crisafulli, "Surge tank control in a cane raw sugar factory," *J. Process Control*, no. 9, pp. 33–39, 1999.

[4] H. Witsenhausen, "A class of hybrid-state continuous-time dynamic systems," *IEEE Trans. Automat. Contr.*, vol. 11, no. 2, pp. 161–167, Apr. 1966.

[5] D. Liberzon, *Switching in systems and control*. Boston: Birkhäuser Boston Inc, 2003.

[6] R. Goebel, R. Sanfelice, and A. Teel, "Hybrid dynamical systems," *Control Syst. IEEE*, vol. 29, no. 2, pp. 28–93, 2009.

[7] J. Lygeros, K. H. Johansson, S. N. Simic, Jun Zhang, and S. S. Sastry, "Dynamical properties of hybrid automata," *IEEE Trans. Automat. Contr.*, vol. 48, no. 1, pp. 2–17, Jan. 2003.

[8] J. Lunze and D. Lehmann, "A state-feedback approach to event-based control," *Automatica*, vol. 46, no. 1, pp. 211–215, Jan. 2010.

[9] S. Di Cairano, A. Bemporad, and J. Júlvez, "Event-driven optimization-based control of hybrid systems with integral continuous-time dynamics," *Automatica*, vol. 45, no. 5, pp. 1243–1251, May 2009.

[10] X. Xu and P. J. Antsaklis, "Results and perspectives on computational methods for optimal control of switched systems," 2003, pp. 540–555.

[11] F. Borrelli, M. Baotić, A. Bemporad, and M. Morari, "Dynamic programming for constrained optimal control of discrete-time linear hybrid systems," *Automatica*, vol. 41, no. 10, pp. 1709–1721, 2005.

[12] B. Lincoln and A. Rantzer, "Relaxing dynamic programming," *IEEE Trans. Automat. Contr.*, vol. 51, no. 8, pp. 1249–1260, Aug. 2006.

[13] A. Rantzer, "Relaxed dynamic programming in switching systems," *Control Theory Appl. IEE Proc.* -, vol. 153, no. 5, pp. 567–574, Sep. 2006.

[14] A. Bemporad and M. Morari, "Control of systems integrating logic, dynamics, and constraints," *Automatica*, vol. 35, pp. 407--427, 1999.

[15] T. R. Mehta and M. Egerstedt, "An optimal control approach to mode generation in hybrid systems," *Nonlinear Anal.*, vol. 65, no. 5, pp. 963–983, Sep. 2006.

[16] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT Press, 1998.

[17] M. P. Deisenroth, C. E. Rasmussen, and J. Peters, "Gaussian process dynamic programming," *Neurocomputing*, vol. 72, no. 7–9, pp. 1508–1524, Mar. 2009.

[18] L. Busoniu, R. Babuska, B. De Schutter, and D. Ernst, *Reinforcement learning and dynamic programming using function approximators*, 1st ed. CRC Press, 2010.

[19] T. R. Mehta and M. Egerstedt, "Multi-modal control using adaptive motion description languages," *Automatica*, vol. 44, no. 7, pp. 1912–1917, Jul. 2008.

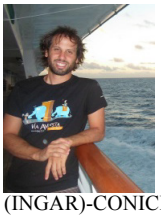
[20] T. Mehta and M. Egerstedt, "Learning multi-modal control programs," vol. 3414, M. Morari and L. Thiele, Eds. Springer Berlin / Heidelberg, 2005, pp. 466–479.

[21] K. J. Astrom and B. M. Bernhardsson, "Comparison of Riemann and Lebesgue sampling for first order stochastic systems," vol. 2, pp. 2011–2016, 2002.

[22] I. Simeonova, "On-line periodic scheduling of hybrid chemical plants with parallel production lines and shared resources," Belgium, 2008.

[23] S. Geist, D. Gromov, and J. Raisch, "Timed discrete event control of parallel production lines with continuous outputs," *Discret. Event Dyn. Syst.*, vol. 18, no. 2, pp. 241–262, Jun. 2008.

[24] G. E. Monahan, "State of the art - A survey of partially observable Markov Decision Processes: theory, models, and algorithms," *Manage. Sci.*, vol. 28, no. 1, pp. 1–16, Jan. 1982.



Mariano De Paula nació en Buenos Aires, y se recibió de Ingeniero Industrial en la Universidad Nacional del Centro de la Provincia de Buenos Aires (UNICEN), Argentina. Luego fue becario del Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) y obtuvo su Doctorado en Ingeniería Industrial de la Universidad Tecnológica Nacional (UTN) realizando sus investigaciones en el Instituto de Desarrollo y Diseño (INGAR)-CONICET. Posteriormente realizó un Posdoctorado en el Centro de Instigaciones en Física e Ingeniería del Centro de la Provincia de Buenos Aires, dependiente del CONICET. Actualmente se desempeña como Investigador del CONICET en dicho Centro y como Profesor en la UNICEN, Argentina. Su área de investigación es la optimización y el control adaptivo mediante técnicas de inteligencia artificial.

Carlos R. Sanchez Reinoso nació en Argentina, en 1981 y recibió el grado de Ingeniero Electrónico con honores por la Facultad de Tecnología y Ciencias Aplicadas de la Universidad Nacional de Catamarca, Argentina, en 2006. Desde 2007 pertenece al Departamento de Electrónica de la Facultad de Tecnología y Ciencias Aplicadas de la Universidad Nacional de Catamarca. Fue becario de investigación del Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina y trabajó en el Instituto de Señales, Sistemas e Inteligencia Computacional de la Universidad Nacional del Litoral (UNL)-CONICET, donde realizó su Doctorado en Ingeniería. También realizó un Posdoctorado en el Instituto de Energía Eléctrica de la Universidad Nacional de San Juan, Argentina. Realizó estancias de investigación en el Centro de Investigación y Estudios Avanzados (CINVESTAV), México y la NASA (USA). Además, fue Profesor Visitante en la Stanford University y Berkeley University, USA. Actualmente es Investigador del CONICET y Profesor Full en la Universidad Nacional de Catamarca, Argentina. Sus intereses de investigación incluyen el modelado matemático, optimización, inteligencia artificial y control con aplicaciones en varios campos.

Luis O. Avila nació en San Luis, Argentina. Recibió el grado de Ingeniero Electrónico por parte de la Universidad Nacional de San Luis (UNSL) y posteriormente su Doctorado en Ingeniería en la Universidad Tecnológica Nacional (UTN), Argentina. Actualmente es becario Posdoctoral del CONICET y desarrolla sus investigaciones en el Laboratorio de Investigación y Desarrollo en Inteligencia Computacional de la UNSL. Sus intereses de investigación son el control y la inteligencia artificial aplicados.