

APRENDIZAJE INDIRECTO CON RETROPROPAGACIÓN EN MODELO NEURONAL DE CAJA GRIS: APLICACIÓN A FERMENTACIÓN EN SUSTRATO SÓLIDO

FRANCISCO CRUZ NARANJO, GONZALO ACUÑA LEIVA

*Departamento de Ingeniería Informática, Universidad de Santiago de Chile
Av. Ecuador 3659, Santiago, CHILE
E-mails: francisco928@gmx.net , gacuna@usach.cl*

FRANCISCO CUBILLOS

*Departamento de Ingeniería Química, Universidad de Santiago de Chile
Casilla 10233, Santiago, CHILE
E-mail: fcubillo@lauca.usach.cl*

Abstract— Grey-Box neural models mix differential equations which act as white box and neural networks, used as black box to complete the phenomenological model. These models have been used in different researches proving their efficacy. The aim of this work is to show the training of the Grey-Box model through indirect backpropagation and Levenberg-Marquardt in Matlab[®]. The model was proved in a batch biotechnological process simulation of solid substrate fermentation, responding to the simulation successfully.

Keywords— Grey-Box neural model, neural networks, identifiability, time-varying parameters, biotechnological processes.

Resumen— Los modelos neuronales de caja gris mezclan ecuaciones diferenciales, las cuales actúan como caja blanca y redes neuronales, utilizadas como caja negra para complementar el modelo fenomenológico. Estos modelos han sido utilizados en diversas investigaciones probándose su eficacia. El presente trabajo tiene por objetivo mostrar el entrenamiento del modelo de caja gris mediante retropropagación indirecta y Levenberg-Marquardt en Matlab[®]. El modelo fue probado en la simulación de un proceso biotecnológico batch de fermentación en sustrato sólido, respondiendo satisfactoriamente a la simulación.

Palabras-clave— Modelo neuronal de caja gris, redes neuronales, identificación, parámetros variantes en el tiempo, procesos biotecnológicos.

1 Introducción

Los bioprocesos resultan altamente interesantes para el estudio de nuevas tecnologías, es por esto que surge la motivación de monitorear y optimizar tales procesos. Sin embargo al tratarse de procesos complejos, como aquellos en los que intervienen gran cantidad de variables de entrada y salida y que además son representados por modelos no lineales y con parámetros variables en el tiempo, surgen importantes impedimentos.

Cuando los procesos a modelar son complejos, la determinación de variables o parámetros relevantes para el mejoramiento de éstos es una tarea ardua y difícil. Ante esto surge la necesidad de estimar las variables que no se pueden medir directamente. Para esto se requiere un sensor virtual, el cual observa y mide directamente otras variables del sistema, con lo cual puede estimar las variables que no pueden ser medidas en línea (James et. al, 2002).

Un problema adicional constituye un modelo que posee parámetros variantes en el tiempo, ya que se debe adoptar una estrategia para identificar tales parámetros en línea y tiempo real. Una metodología utilizada en estos casos, especialmente en el ámbito de procesos químicos y biotecnológicos, son los denominados modelos de caja gris (Hensen et. al,

2000). Estos modelos son aquellos que incluyen un modelo fenomenológico limitado, el cual se complementa con parámetros obtenidos mediante redes neuronales.

Por otra parte, uno de los factores de mayor relevancia en un proceso biotecnológico es la concentración de biomasa, que es definida como la masa o número de células, o como el total de masa microbiana. En los procesos biotecnológicos se busca la producción de altas concentraciones de biomasa o producto (Aguilar & Filho, 2001).

Los bioprocesos son afectados por una gran cantidad de variables, las cuales influyen en forma directa el metabolismo de las células, muchas de las cuales pueden ser medidas directamente en línea, como por ejemplo el pH o la temperatura. Otras variables, como la biomasa, no pueden ser medidas en línea (Chen et. al, 2000), muchas veces por no existir el instrumento capaz de hacer la medición, otras por el elevado costo, o bien porque la medición podría tardar varios días, siendo el objetivo lograr una medición inmediata.

Las estrategias de aprendizaje o entrenamiento utilizadas hasta ahora suponen la existencia de datos para los parámetros obtenidos por el modelo neuronal (James et. al, 2002), sin embargo esto no es posible la mayor parte de las veces.

El aprendizaje utilizado en este trabajo no necesita datos para la parte neuronal, trabajando de esta

manera solo con los datos que ve la parte fenomenológica del modelo lo que se detallará más adelante. La creación del modelo propuesto, el aprendizaje y las simulaciones fueron realizadas totalmente en ambiente Matlab®, efectuándose simulaciones del tipo sensor virtual.

2 Modelos de Caja Gris

Los modelos neuronales de caja gris son utilizados para sistemas donde existe algo de conocimiento a priori, es decir, se conocen algunas leyes físicas, pero sin embargo, algunos parámetros se deben determinar a partir de datos observados.

Los modelos de caja gris, inicialmente denominados redes neuronales híbridas o modelos híbridos neuronales, han sido estudiados hace más de una década. En (Psichogios & Ungar, 1992) se desarrollan fundamentos de una primera red neuronal híbrida para modelado de procesos. En su estudio realizaron una comparación entre redes neuronales estándar y redes neuronales híbridas resultando éstas últimos con mejor desempeño al predecir parámetros, debido a poseer conocimiento a priori y el correspondiente aprovechamiento de éste. Además probaron que los modelos híbridos pueden ser identificados y entrenados con un grupo bastante menor de datos, con respecto a un modelo equivalente de caja negra (redes neuronales).

Dos años más tarde, en (Thompson & Kramer, 1994) se clasificaron los modelos híbridos en dos categorías. La primera es el modelo de caja gris en serie (Fig. 1), el cual consistente en una red neuronal que soporta valores intermedios, que luego son introducidos en el modelo fenomenológico. La segunda categoría es el modelo de caja gris en paralelo (Fig. 2), el cual consistente en una red neuronal que compensa el modelo fenomenológico, en el sentido de modelar el error.

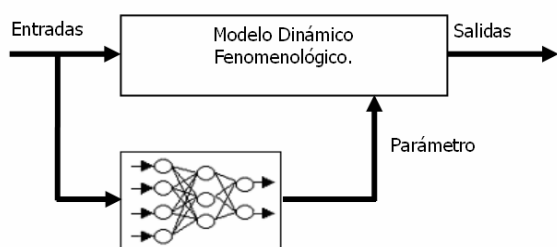


Figura 1. Modelo neuronal de caja gris en serie.

Luego en (Van Can et. Al, 1996) se demostró que en general el esquema híbrido de caja gris en serie presenta mejores resultados, comparativamente con el esquema en paralelo, mostrando mejores rendimientos de modelos en serie para sistemas MISO (del inglés, multiple input single output). No obstante en modelos para sistemas SISO (del inglés, single input single output) los resultados mostrados son de características similares.

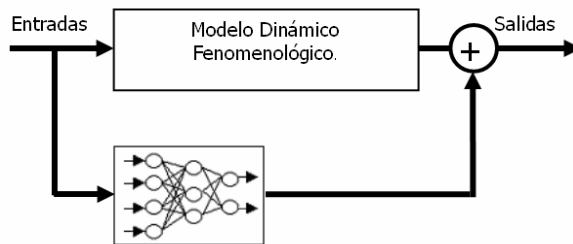


Figura 2. Modelo neuronal de caja gris en paralelo.

En (Acuña et. al, 1999), se distinguen entre dos formas de entrenamiento. La primera forma corresponde al entrenamiento directo (Fig. 3), la cual utiliza el error originado a la salida de la red neuronal para la correcta determinación de sus pesos. La segunda forma es el entrenamiento indirecto (Fig. 4), en donde se utiliza el error originado a la salida del modelo para efectos de aprendizaje de la red neuronal. El entrenamiento indirecto se puede realizar de dos formas, la primera es minimizando una función objetivo de variables de estado, por medio de una técnica de optimización no lineal y la segunda es retropropagando el error de salida en los pesos de la red neuronal a través de las ecuaciones del modelo fenomenológico.

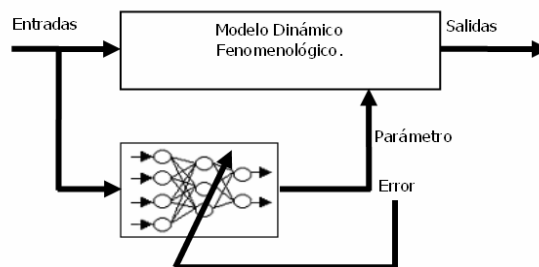


Figura 3. Modelo de caja gris con entrenamiento directo

En (Pinto & Acuña, 2004), se desarrolla una aplicación basada en Matlab® para la elaboración de modelos neuronales de caja gris utilizando una estrategia de entrenamiento directa, este trabajo tiene por objetivo realizar una implementación en Matlab®, de un algoritmo de entrenamiento basado en Levenberg-Marquardt con estrategia indirecta, retropropagando el error del modelo a través del modelo fenomenológico hacia la red neuronal.

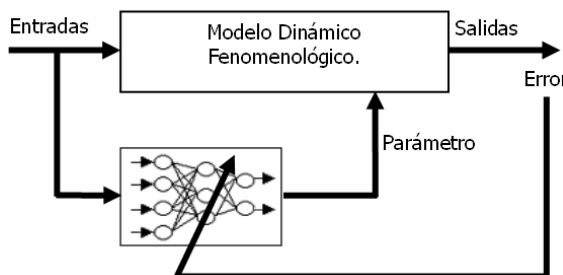


Figura 4. Modelo de caja gris con entrenamiento indirecto

3 Proceso Biotecnológico

En este trabajo se propone un modelo neuronal de caja gris para la simulación de un bioproceso batch de fermentación sobre sustrato sólido, correspondiente a la producción de ácido giberélico, obtenido desde el hongo *Gibberella Fujikuroi*.

La *Gibberella Fujikuroi* es utilizada para producir sustancias naturales llamadas giberelinas o ácido giberélico, éstos son compuestos que actúan como hormonas vegetales, utilizados para mejorar tamaño, color y aspecto general de algunas especies vegetales, tales como naranjas, uvas, manzanas, tomates, entre otras. Por otra parte, en la industria cervecera son de gran utilidad al disminuir el tiempo de producción de la malta en forma considerable. En la agroindustria su uso, principalmente, es relacionado con la maduración y crecimiento de la uva de mesa sin semilla.

Por otra parte, un cultivo batch es un proceso en que las células crecen en un sistema cerrado, sin modificar sus condiciones, hasta que el proceso de crecimiento es limitado por la cantidad de nutriente. El cultivo sobre sustrato sólido es referido al crecimiento microbiano y a los metabolitos producidos en el interior de un medio sólido y bajo condiciones de humedad restringida.

Un modelo simplificado describe la evolución de las principales variables (Gelmi et al, 2001). Este modelo fenomenológico basado en las leyes de variable de masa considera 8 variables de estado: biomasa activa (X), biomasa medida (X_m), urea (U), nitrógeno intermedio (N_I), almidón (S), ácido giberélico (GA₃), dióxido de carbono (CO₂) y oxígeno (O₂). Solo las dos últimas variables pueden ser directamente medidas en línea. Las ecuaciones del modelo discretas con el método de Euler en el intervalo t y t+1 son las siguientes:

$$X_{measu(t+1)} = X_{measu(t)} + \left(\mu \cdot X(t) \right) \cdot \Delta t \quad (1)$$

$$X_{(t+1)} = X(t) + \left(\mu \cdot X(t) - k_d \cdot X(t) \right) \cdot \Delta t \quad (2)$$

$$U_{(t+1)} = U(t) + (-k) \cdot \Delta t \quad (3)$$

$$N_{I(t+1)} = \begin{cases} N_{I(t)} + \left(0,47 \cdot k - \mu \cdot \left(\frac{X(t)}{Y_{X/N_I}} \right) \right) \cdot \Delta t \\ \text{si } U \geq 0 \\ N_{I(t)} + \left(-\mu \cdot \left(\frac{X(t)}{Y_{X/N_I}} \right) \right) \cdot \Delta t \text{ si } U < 0 \end{cases} \quad (4)$$

$$S_{(t+1)} = S(t) + \left(-\frac{\mu \cdot X(t)}{Y_{X/S}} - m_s \cdot X(t) \right) \cdot \Delta t \quad (5)$$

$$GA_{3(t+1)} = GA_{3(t)} + \left(\beta \cdot X(t) - k_p \cdot GA_{3(t)} \right) \cdot \Delta t \quad (6)$$

$$CO_{2(t+1)} = CO_{2(t)} + \left(\mu \cdot \left(\frac{X(t)}{Y_{X/CO_2}} \right) + m_{CO_2} \cdot X(t) \right) \cdot \Delta t \quad (7)$$

$$O_{2(t+1)} = O_{2(t)} + \left(\mu \cdot \left(\frac{X(t)}{Y_{X/O_2}} \right) + m_{O_2} \cdot X(t) \right) \cdot \Delta t \quad (8)$$

Las salidas medidas de modelos son las siguientes:

$$y_1 = CO_{2(t+1)} \quad (9)$$

$$y_2 = O_{2(t+1)} \quad (10)$$

Por otra parte los parámetros difíciles de obtener y que serán estimados por la parte neuronal del modelo son μ y β , correspondientes a la tasa de crecimiento específica y a la tasa de producción específica de ácido giberélico, respectivamente, cuyas ecuaciones, utilizadas en la simulación para la generación de datos, son las siguientes:

$$\mu = \frac{\mu_{max} \cdot N_I}{(N_I + k_n)} \quad (11)$$

$$\beta = \frac{\beta_{etam}}{1 + K_i \cdot N_I} \quad (12)$$

Los otros parámetros del modelo fueron identificados en base a prácticas específicas y condiciones experimentales. Sus valores son incluidos en la Tabla 1 en condiciones de temperatura y actividad de agua controlada (T=25°C, Aw=0,992).

4 Solución Propuesta

La solución propuesta es un modelo neuronal de caja gris, en donde, la parte fenomenológica de éste puede ser descrita en conjunto con la parte empírica del mismo, para lo cual es diseñada una red neuronal que contenga ambas partes del modelo de caja gris (Fig. 5). Esta red neuronal híbrida posee la capacidad de fijar pesos en la fase de entrenamiento, para así actuar como un modelo de caja gris. Los pesos en la Fig. 5 que poseen un valor señalado corresponden a la parte fenomenológica del modelo, de esta manera, estos pesos no serán modificados durante el entrenamiento. Los pesos en los que no se indica valor corresponden a la parte neuronal del modelo. Estos pesos, inicialmente, fueron asignados con valores pseudoaleatorios obtenidos con el método de inicialización de (Nguyen & Widrow, 1990).

Se puede observar en la Fig. 5 que uno de los pesos correspondiente a la parte de caja blanca o

fenomenológica está graficado mediante una línea punteada, esta línea representa el cambio de modelo que se aprecia en la cuarta variable de estado (Ni) en el modelo matemático, es decir, si la Urea (U) es mayor o igual que cero este peso tiene el valor señalado, de lo contrario, si la Urea (U) es menor que cero, este peso tiene valor cero.

Tabla 1. Parámetros del modelo

Nom-bre	Descripción	Valor [Unidad]
μ_m	Máxima tasa de crecimiento específica	0.28 [1/h]
β_{etam}	Coefficiente del sustrato de inhibición	$6.5 \cdot 10^{-4}$ [g GA ₃ /g X h]
k	Tasa de conversión de urea a nitrógeno disponible	$1.33 \cdot 10^{-4}$ [g/h]
k_n	Constante de "Michaelis-Menten"	$1.1 \cdot 10^{-3}$ [g N/g.i.s.]
k_d	Tasa de mortalidad de la biomasa	0.031 [1/h]
k_p	Tasa de degradación del ácido giberélico	$4.4 \cdot 10^{-4}$ [1/h]
k_i	Tasa de producción del nitrógeno intermedio	$7.86 \cdot 10^5$ [1/g N/g.i.s.]
m_s	Coefficiente de mantenimiento por almidón	0.11 [g S / g X h]
m_{CO_2}	Coefficiente de mantenimiento por CO ₂	0.11 [g CO ₂ / g X h]
m_{O_2}	Coefficiente de mantenimiento por O ₂	0.06 [g O ₂ / g X h]
$Y_{X/NI}$	Coefficiente de rendimiento biomasa / nitrógeno intermedio	20.8 [g X / g N _I]
$Y_{X/S}$	Coefficiente de rendimiento biomasa / almidón	1.21 [g X / g S]
Y_{X/CO_2}	Coefficiente de rendimiento biomasa / CO ₂	0.58 [g X / g CO ₂]
Y_{X/O_2}	Coefficiente de rendimiento biomasa / O ₂	2.11 [g X / g O ₂]

En la Fig. 5 se puede distinguir claramente que la parte neuronal del modelo es estimada por un perceptron multicapa, que se encuentra inserto en el modelo neuronal de caja gris, el cual posee una arquitectura de 1 x 6 x 2. Su entrada corresponde sólo a Ni debido a que como se observó en el modelo matemático los dos parámetros difíciles de obtener tienen dependencia solo con esta variable de estado.

En consecuencia, el perceptron multicapa estima los valores de los dos parámetros difíciles de obtener, los cuales se mezclan a su vez con la parte fenomenológica del modelo, para así obtener la salida de éste.

Para la parte neuronal del modelo se utilizó tangente hiperbólica como función de transferencia en la capa intermedia y función identidad en la capa de salida, mientras que para la parte fenomenológica se utilizó la función identidad como función de transfe-

rencia. La función de activación utilizada fue la sumatoria de las ponderaciones de las entradas por los pesos, exceptuando las dos neuronas inmediatamente después de la salida de la parte neuronal, o bien la capa intermedia de la red, donde se utilizó una productoria para la ponderación de las entradas por los pesos, esto debido a la forma de las ecuaciones fenomenológicas.

Para la validación del modelo neuronal de caja gris propuesto se realiza el cálculo de índices de calidad, tales como el IA, RMS y RSD, los valores considerados aceptables para estos índices son IA>0.9, RMS <0.1 y RSD<0.1. Las ecuaciones de los índices de calidad son las siguientes:

$$IA = 1 - \frac{\sum_{i=1}^n (o_i - p_i)^2}{\sum_{i=1}^n (|o_i| + |p_i|)^2} \quad RMS = \sqrt{\frac{\sum_{i=1}^n (o_i - p_i)^2}{\sum_{i=1}^n o_i^2}} \quad RSD = \sqrt{\frac{\sum_{i=1}^n (o_i - p_i)^2}{N}}$$

Donde o_i y p_i son los valores observados y predichos respectivamente en el tiempo i , y N es el número total de datos. Luego, $p_i' = p_i - o_m$ y $o_i' = o_i - o_m$, donde o_m es el valor medio de las observaciones.

5 Entrenamiento

En la red se utiliza como función de *performance* el error cuadrático medio, que es calculado en la salida del modelo fenomenológico y desde ahí retropropagado hacia los pesos de la red que son modificables. En la fase de entrenamiento se utilizaron 70% de los datos disponibles y los restantes fueron utilizados para validación.

El algoritmo de entrenamiento corresponde al método de Levenberg-Marquardt Backpropagation que es un algoritmo de segundo orden que presenta una ligera modificación sobre el método tradicional de Newton. Además el algoritmo posee la capacidad de modificar sólo los pesos que le sean señalados, dejando así un grupo de pesos fijos que, como ya se mencionó, representan la parte fenomenológica del modelo, en la fase de entrenamiento.

Este esquema de entrenamiento corresponde al segundo método indirecto de entrenamiento propuesto por (Acuña et. Al, 1999), en el cual se calcula el error en la salida del modelo de caja gris y desde ahí retropropagado hacia la parte empírica o neuronal del modelo. Esto cuenta con la ventaja que no es necesario poseer mediciones de los parámetros difíciles de obtener.

La red neuronal híbrida o modelo de caja gris fue entrenada por 100 épocas, realizándose en varias oportunidades, obteniéndose alrededor de mil entrenamientos distintos, para luego seleccionar la mejor de éstas, utilizando como criterio de selección el error cuadrático medio obtenido a la salida del modelo. Posteriormente en la fase de simulación se calcularán los índices error IA, RMS y RSD. Es importante notar que el número de épocas de entrenamiento

puede aumentar en este tipo de modelos, ya que al poseer pesos fijos es una restricción que se añade al algoritmo de optimización.

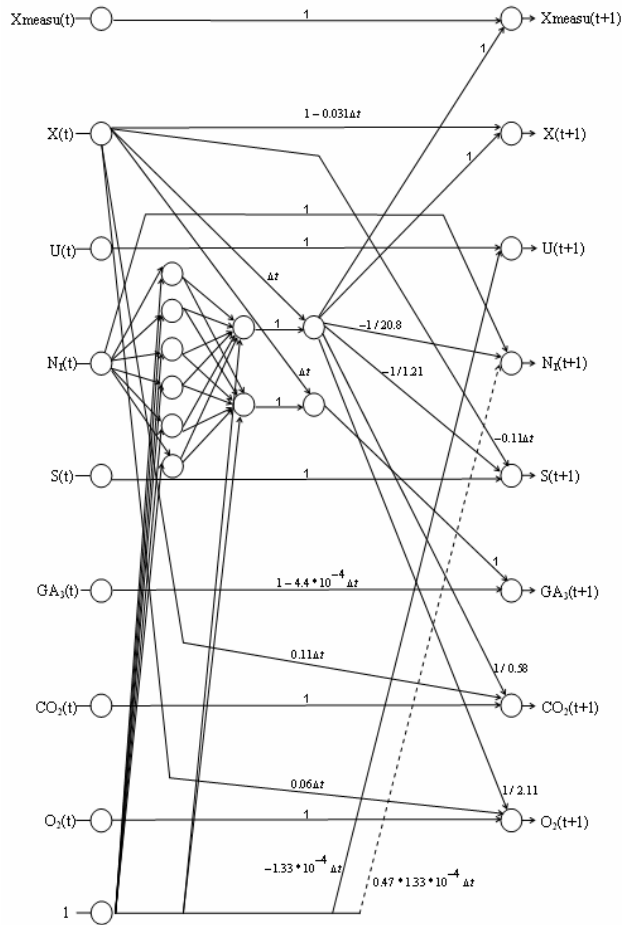


Figura 5. Modelo de caja gris para proceso de fermentación de sustrato sólido.

6 Simulación y Resultados

Para la simulación se realizaron pruebas para datos sin error, con un 5% de error y con un valor inicial erróneo para la biomasa. El esquema de simulación del sensor virtual se muestra en la Fig. 6. La simulación se realizó para los mil datos disponibles.

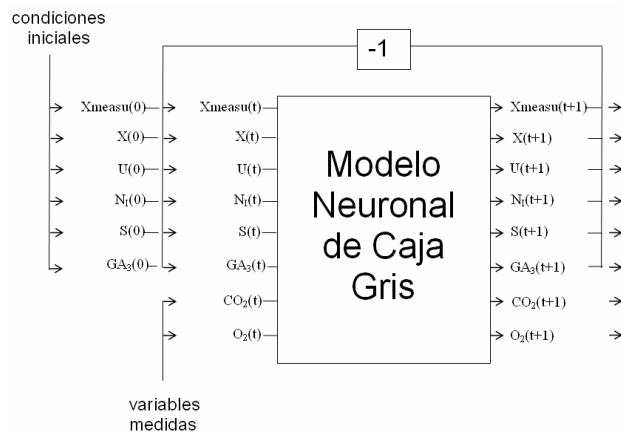


Figura 6. Esquema de simulación del sensor virtual.

Las condiciones iniciales en operación normal son las siguientes:

$$X(0) = \begin{bmatrix} 0 & 0.0040 & 0.0040 & 0.5 * 10^{-4} & 0.0040 & 0 & 0 & 0 \end{bmatrix}$$

Caso 1: simulación sin ruido

El primer caso evaluado fue el caso de los datos sin ruido para verificar la validez del estudio, la Fig. 7 muestra el valor real y estimado de biomasa, nótese que ambas líneas están superpuestas. Los índices de calidad obtenidos fueron de IA=0.99, RMS=0.2E-3 y RSD=0.32E-1.

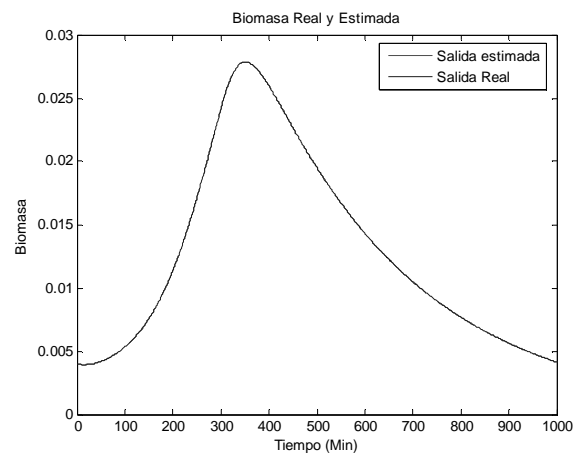


Figura 7. Biomasa real y estimada para el caso 1

Caso 2: simulación con ruido

El segundo caso evaluado corresponde a la simulación con un 5% de ruido en los datos de entrada. La Fig. 8 muestra la biomasa real y estimada. En este caso los índices de calidad obtenidos fueron los siguientes: IA=0.99, RMS=0.5E-1 y RSD=0.8E-3.

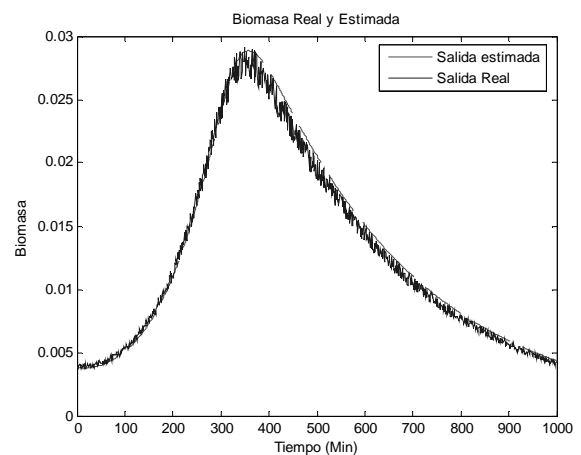


Figura 8. Biomasa real y estimada para el caso 2

Caso 3: simulación con ruido en los datos y condición inicial errónea en la biomasa

El último caso evaluado corresponde a la simulación con un 5% de ruido en los datos de entrada y además una condición inicial distinta para la biomasa con un error del 250%, para verificar la convergencia del modelo. La Fig. 9 muestra la respuesta del modelo, ahí se aprecia la biomasa real y la estimada. Los índices de calidad fueron $IA=0.93$, $RMS=0.24$ y $RSD=0.41E-2$. En este caso el índice RMS no cumple con la condición aceptable ($RMS<0.1$), sin embargo esto se debe a la naturaleza del error introducido, ya que el modelo tarda alrededor de 300 iteraciones para ajustar la curva estimada al valor real. Esta simulación, sin embargo nos muestra la convergencia del sensor virtual basado en el modelo de caja gris, por lo que se considera aceptable.

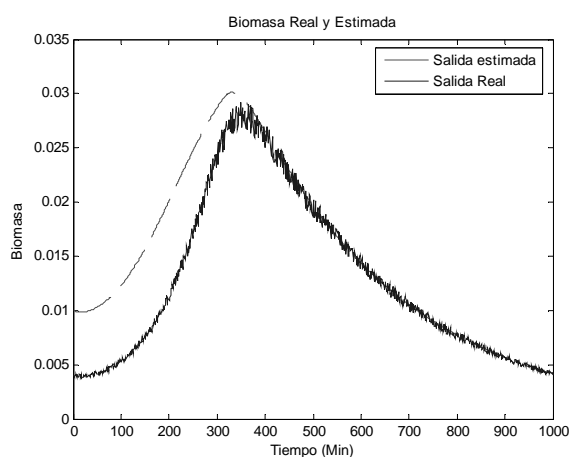


Figura 9. Biomasa real y estimada para el caso 3

Conclusiones

Los modelos neuronales de caja gris son una real alternativa para modelar procesos del mundo real. Poseen ventajas sobre los modelos de caja negra, ya que se apoyan en el conocimiento a priori que se tenga sobre proceso.

El modelo planteado en este trabajo combina el conocimiento fenomenológico del proceso con una red neuronal de tipo perceptrón multicapa, para la estimación de parámetros dentro de una red neuronal de orden mayor para llevar a cabo el proceso de retropropagación.

El método de entrenamiento indirecto, con retropropagación y Levenberg-Marquard, utilizado en este trabajo muestra buenos resultados en la fase de aprendizaje y posterior simulación sólo con las variables medibles a manera de sensor virtual.

Agradecimientos

Los autores desean agradecer el financiamiento parcial del Proyecto Fondecyt 1040208.

Referencias Bibliográficas

- Acuña, G., Cubillos, F., Thibault, J., Latrille, E. (1999). *Comparison of methods for training grey-box neural network models*, Computers and Chemicals Engineering Supplement, pp. 561-564.
- Aguiar, H., Filho, R. (2001). *Neural network and hybrid model: a discussion about different modeling techniques to predict pulping degree with industrial data*. Chemical Engineering Science 56, pp. 565-570.
- Chen, L., Bernard, O., Bastin, G., Angelov, P. (2000). *Hybrid modelling of biotechnological processes using neural networks*. Control Engineering Practice 8, pp. 821-827.
- Gelmi, C., Pérez-Correa, R., Agosin, E. (2002). *Modelling Gibberella fujikuroi growth and GA₃ production in solid-state fermentation*. Process Biochemistry 37, pp. 1033-1040.
- Hensen, R., Angelis, G., van de Molengraft, M., de Jager, A., Kok, J. (2000). *Grey-box modeling of friction: An experimental case-study*. European Journal of Control 6, pp. 258-267.
- James, S., Legge, R., Budman, H. (2000). *Comparative study of black-box and hybrid estimation methods in fed-batch fermentation*. Journal of Process Control 12, pp. 113-121.
- Nguyen, D., Widrow, B. (1990). *Improving the Learning Speed of 2-Layer Neural Networks by Choosing Initial Values of the Adaptive Weights*. Proceedings IEEE IJCNN, pp. 21-26, San Diego.
- Pinto, E., Acuña, G. (2004). *Desarrollo de un Toolbox Matlab para la elaboración de modelos neuronales de caja gris*. Actas en CD del XVI Congreso de la Asociación Chilena de Control Automático, Universidad de las Américas, Santiago, Noviembre 2004.
- Psichogios, D., Ungar, L. (1992). *A Hybrid Neural Network-First Principles Approach to Process Modeling*, Vol. 38, N°10, pp. 1499-1511.
- Thompson, M.; Kramer, M. (1994). "Modeling Chemical Processes Using Prior Knowledge and Neural Networks", Vol. 40, N°8, páginas 1328-1340.
- Van Can, H., Hellinga, C., Luyben, K., Heijnen, J. (1996). *Strategy for Dynamic Process Modeling Based on Neural Network in Macroscopic Balance*", AIChE 42, pp. 3403-3418.