

LAS REDES NEURONALES ARTIFICIALES EN QUÍMICA ANALÍTICA. PARTE I. FUNDAMENTOS

EDWINA. HERNÁNDEZ-CARABALLO* - RITAMa. AVILA. G
FRANKLIN RIVAS

Resumen

La cantidad de datos que pueden generar los cada vez más sofisticados equipos de análisis instrumental, así como el tiempo que puede tomar la interpretación de los mismos, pueden a menudo abrumar a cualquier químico analítico. Es por ello que se hace necesario el empleo de herramientas para análisis y procesamiento de datos igualmente poderosas. Las redes neuronales artificiales multi-capas entrenadas con el algoritmo de retro-propagación, se cuentan entre el grupo de las llamadas técnicas de computación inteligente disponibles para tal fin. El propósito del presente trabajo es discutir, de forma general, los fundamentos y capacidades del tipo de red neuronal artificial previamente mencionada, así como los problemas asociados al uso de la misma.

Abstract

The amount of data that can be obtained from the increasingly sophisticated analytical instrumentation, and hence the time involved in its interpretation may often overwhelm the analytical chemist. It is therefore necessary the use of equally powerful analysis and processing tools. Multi-layer artificial neural networks trained with the backpropagation algorithm are among the so-called intelligent computation techniques available for such purpose. The aim of this work is presenting an overview of the foundations and capacities of the aforementioned type of artificial neural network, as well as the problems associated with its use.

Introducción

El químico analítico dispone actualmente de una enorme variedad de técnicas instrumentales que le permiten atacar una cantidad de problemas de las más diversas índoles. La microelectrónica y la computación, a su vez, han conducido a un incremento en la cantidad de datos que se pueden obtener en cada experimento. Ejemplo de ello lo constituyen los voltagramas/polarogramas, los espectros infrarrojos de compuestos individuales, o los espectros UV-visible de mezclas, entre otros. La extracción de la información útil a partir de dichos datos hace necesaria la disponibilidad de técnicas computacionales

igualmente poderosas. Convencionalmente, esto se ha llevado a cabo a través de la aplicación de técnicas *quimiométricas*, e. g., análisis por componentes principales, análisis discriminante, etc.¹ Más recientemente, y con el mismo fin, se han venido aplicando una serie de estrategias que se agrupan dentro de lo que se conoce como *técnicas de computación inteligente*,² entre las que se destacan las *redes neuronales artificiales*.³

El presente trabajo está estructurado en dos partes. La primera de ellas está dirigida a mostrar, de una forma sencilla, los principios sobre los que se basan las redes neuronales artificiales, así como sus potenciales aplicaciones y los problemas asociados a su empleo.

Es importante destacar que actualmente existe una amplia variedad de redes neuronales y algoritmos de entrenamiento. El presente trabajo, sin embargo, está enfocado a la discusión de un sólo tipo de red neuronal artificial, a saber, la *red multi-capas (o en cascada)* entrenada con el algoritmo de retro-propagación*.

Esta simplificación del tema está motivada principalmente por las siguientes razones:⁴

- Este tipo de red neuronal puede ser utilizada para resolver problemas que involucren *clasificación, representación gráfica, modelado u optimización*, con mínimas modificaciones en su *arquitectura*.
- La simplicidad de su desarrollo, aunado a su versatilidad, han hecho de este tipo de red la más empleada para la resolución de problemas en el área química.

La segunda parte del trabajo, la cual será publicada posteriormente, estará dedicada a la discusión de algunas aplicaciones de las redes neuronales artificiales en la química analítica. Se hará particular énfasis en la sinergia entre estas herramientas y algunas de las técnicas instrumentales más

*Ambas expresiones (*multi-capas o en cascada*) hacen alusión a la forma como "fluyen" los datos a través de la red neuronal, por lo que se emplearán indiferentemente a lo largo de este trabajo

“populares” disponibles en muchos laboratorios de análisis, e. g., cromatografía, espectroscopía atómica, voltametría, etc.

Fundamentos de las redes neuronales artificiales

Las redes neuronal muti-capas entrenadas con el algoritmo de retro-propagación han sido ampliamente descritas en monografías,²⁻⁵ artículos tutoriales,⁶⁻¹⁰ etc. Nuestro fin es el de mostrar una visión inicial de lo que son estas herramientas y, posteriormente, ejemplificar las capacidades de las mismas a través de algunas aplicaciones químicas. El presente trabajo está dirigido principalmente a lectores con escaso o ningún conocimiento en el área. Por esa razón, el empleo de ecuaciones matemáticas ha sido restringido al mínimo necesario para la comprensión de este tópico. Para una discusión más profunda de los temas aquí tratados el lector interesado puede consultar las referencias.⁶⁻¹⁰

La unidad fundamental: La neurona

Las computadoras biológicas (cerebros) están constituidas por un número variable de unidades que, esencialmente, son las responsables de las maravillas de las que son capaces los seres vivos: *las neuronas*. La figura 1(a) muestra un esquema simplificado de uno de los varios tipos de células nerviosas y las partes que la componen.

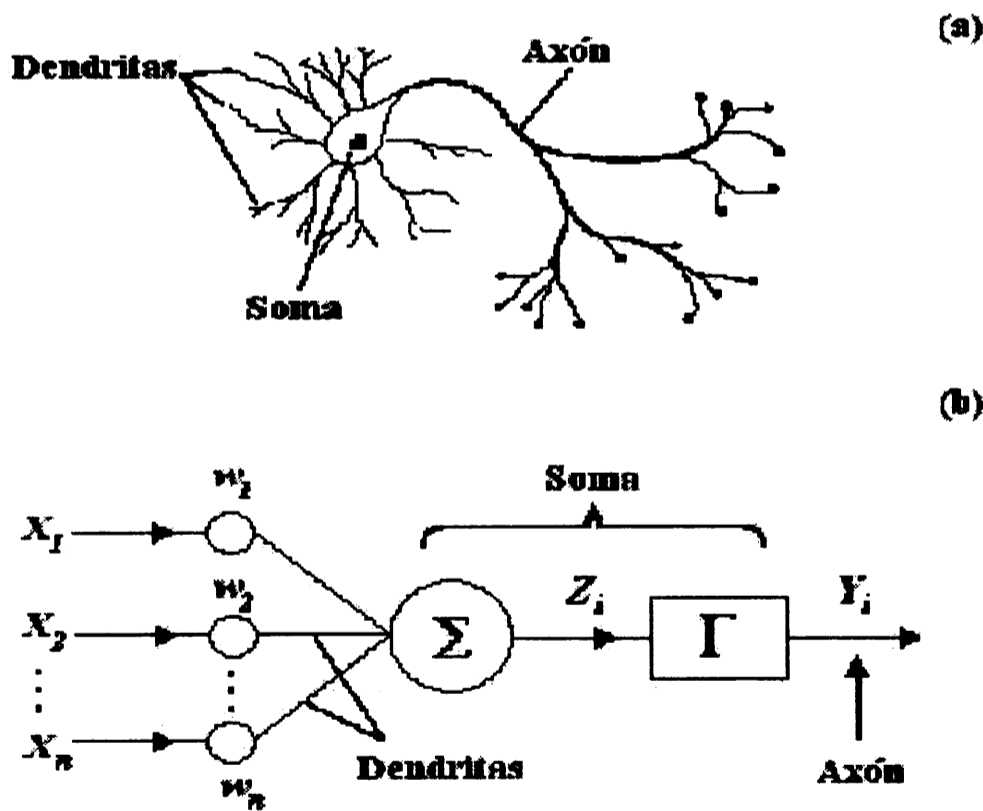


Figura 1. Diagrama simplificado de una neurona biológica (a), y de una artificial (b), con sus principales similitudes.

La función principal de dichas células es la transmisión y el almacenamiento de información necesarias para la adaptación del individuo a su entorno. Resulta lógico entonces que algunos de los intentos llevados a cabo por el hombre para crear *inteligencia artificial*, hayan comenzado con un intento por simular la engañosa simplicidad de dicha unidad. La figura 1(b) muestra un esquema de una neurona artificial.

El procesamiento de los impulsos nerviosos, en el caso biológico, y de los datos, en el caso artificial, es bastante similar, por lo que sólo nos referiremos a cómo esto tiene lugar en esta última instancia. El “cuerpo” de la neurona recibe datos (X_i) de otras unidades a través de múltiples conexiones. La fuerza de dichas entradas es modificada a través de los pesos (w_i), los cuales desempeñan un papel similar al de las *conexiones sinápticas*. Cada neurona artificial lleva a cabo únicamente dos funciones. La primera consiste en una suma ponderada de los datos de entrada:

$$Z_i = \sum_i w_i X_i \quad (1)$$

El valor resultante de esta operación puede ser muy grande o negativo. Desde el punto de vista biológico, esto último no tiene sentido por cuanto se sabe que una neurona emite o no una señal electroquímica (aunque computacionalmente lo anterior no tiene importancia). Es principalmente por la primera razón, i. e., la magnitud de la suma, que el valor resultante es sometido a un segundo procesamiento. Aquí los valores son transformados mediante una *función de transferencia*, la cual reduce el rango de los valores de salida de la neurona a un intervalo más estrecho. La tabla 1 resume algunas de las funciones de transferencia más empleadas en la actualidad, así como los rangos dentro de los cuales son expresados los datos de salida.

Tabla 1. Funciones de transferencias y rangos de los datos de salida.

Nombre	Ecuación	Rango
Sigmoide unipolar	$Y_i = \frac{1}{1 - e^{-(Z_i)}}$	[0,1]
Sigmoide bipolar	$Y_i = \frac{2}{1 - e^{-(Z_i)}} - 1$	[-1,1]
Tangente hiperbólica	$Y_i = \text{Tanh}(Z_i)$	[-1,1]
Lineal	$Y_i = Z_i$	Reales

La figura 2 muestra gráficamente el perfil de una de las funciones de transferencia más frecuentemente empleada: la *sigmoide unipolar*. En dicha Figura se puede apreciar una región en la que la neurona no transmite señal ($Y_i = 0$), independientemente de el valor de entrada ($Z_i < 5$); y una región en la que la señal de salida es máxima ($Y_i = 1$), para valores de entrada $Z_i > 5$. Entre esas dos zonas existe una región de transición en la que la respuesta varía continuamente de manera no lineal (zonas en gris) y/o linealmente (zona en blanco). La capacidad de una neurona (o de una red neuronal) para adaptar

sus pesos de manera de generar una respuesta en alguna de las zonas mencionadas anteriormente, es lo que le otorga a estas herramientas el poder para la ejecución de tan diversas tareas. Lo anterior será visto con más claridad cuando discutamos posteriormente los tipos de problemas que se pueden resolver empleando redes neuronales artificiales.

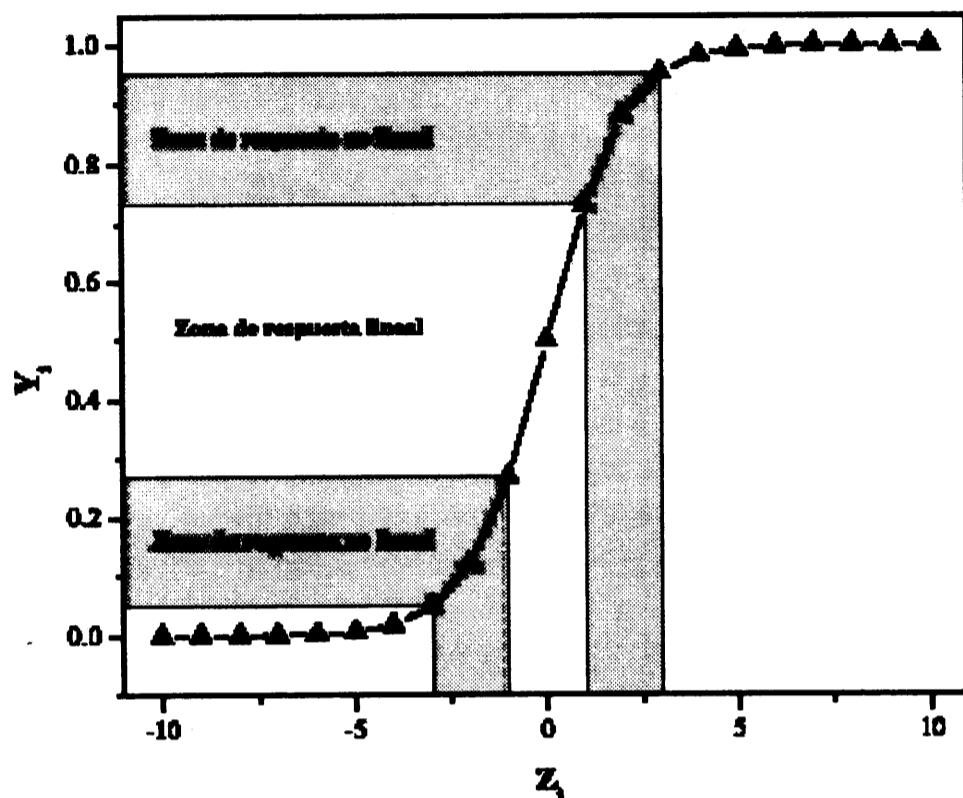


Figura 2. Función de transferencia sigmoideal. El ajuste de los pesos durante la etapa de entrenamiento permite establecer las relaciones adecuadas (lineales o no lineales) entre los datos de salida (Y_i) y los de entrada (Z_i).

La unión hace la fuerza: La red neuronal

Como se vio en la sección anterior, cada neurona lleva a cabo una función muy simple: sumar los "estímulos" y generar una "respuesta". Esto por sí sólo es insuficiente como para explicar la increíble capacidad de los seres vivos para desempeñar multitud de tareas en un medio ambiente cambiante (adaptación). De hecho, algunos investigadores han mostrado que las señales generadas por las neuronas son muy similares entre especies que difieren grandemente en desarrollo (evolutivamente hablando), e. g., un caracol y un profesor universitario.³ Por otra parte, también se sabe que las neuronas son entre cinco y seis órdenes de magnitud más lentas que cualquier procesador de silicón. Sin embargo, un ser humano es capaz de ejecutar algunas tareas a velocidades que no han podido ser emuladas por ningún computador moderno, e. g., reconocer de un vistazo el rostro de un amigo entre una muchedumbre.⁴

De lo anterior se deriva la conclusión de que el poder del cerebro y de su simulación artificial radica no en sus elementos constituyentes *per se*, sino en la unión de ellos: "La unión hace la fuerza".

Una red neuronal artificial no es más que un herramienta computacional diseñada para simular la forma como el cerebro lleva a cabo algunas de sus tareas. La figura 3 muestra un esquema general de una red neuronal artificial del tipo cascada. Este tipo de representación es muy útil en la etapa de diseño de la red, por cuanto permite visualizar fácilmente la forma como fluyen los datos, y como éstos son transformados por cada una de las neuronas que la constituyen. La red neuronal mostrada en la figura 3 consta de cuatro capas activas, i. e., tres capas escondidas y una capa de salida; cada una de las cuales, a su vez, se encuentra conformada por un número variable de neuronas interconectadas entre sí. La capa de entrada sólo sirve para distribuir los datos a todas las neuronas de la capa subsiguiente, por lo que algunos autores no la incluyen al describir la arquitectura* de la red. La capa de salida es la que proporciona la información de vuelta al usuario.

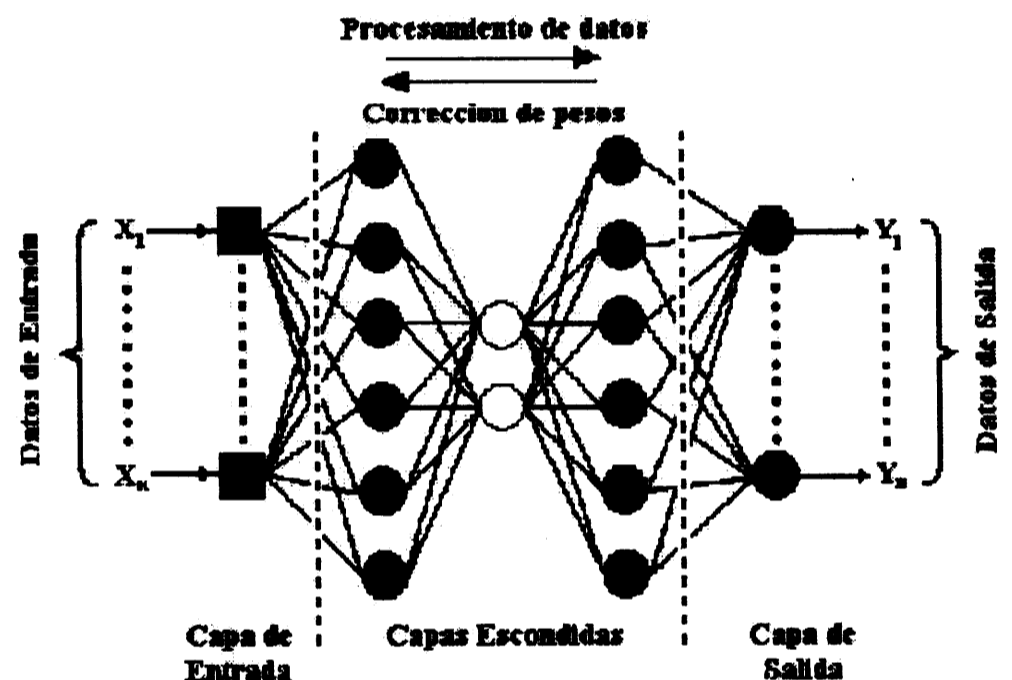


Figura 3. Diagrama simplificado de una red neuronal artificial de tipo cascada. Los datos introducidos por el usuario "fluyen" exclusivamente desde la capa de entrada hacia la capa de salida. La corrección de los pesos al emplear el algoritmo de retro-propagación se lleva a cabo en la dirección contraria (de allí el nombre del algoritmo).

El nombre de "capa(s) escondida(s)" se deriva del hecho que, como se aprecia en la figura, no existe interacción directa entre las neuronas de dicha(s) capa(s) y el usuario. Una restricción que existe en este tipo de arquitectura es que los datos introducidos, así como las transformaciones que sufren los mismos a través de las diversas capas, "viajan" exclusivamente desde la capa de entrada hasta la de salida: no se permiten conexiones entre neuronas de una misma capa o ciclos de retro-alimentación. El ajuste de los pesos, proceso que tiene lugar durante el período de entrenamiento, se lleva a cabo en la dirección contraria: de allí que la estrategia de aprendizaje sea conocida también como *algoritmo de retro-propagación*.

* El lector habrá podido deducir que este término, que fue empleado previamente en la sección introductoria, se refiere al conjunto de características que describen a una red neuronal, i. e., el número de capas, el número de neuronas en cada capa, etc.³

Se ha dicho que una red neuronal con una sola capa escondida es, por sí misma, un “aproximador universal”, es decir, que es capaz de resolver cualquier problema. En este sentido, es el investigador quien debe evaluar y decidir acerca de la necesidad de emplear más de una capa escondida.

El número de *nodos** en la capa de entrada y de neuronas en la capa de salida viene determinado por las características de cada problema, y dependen del número de descriptores y soluciones del mismo. Por el contrario, el número de neuronas y capas escondidas, así como las funciones de transferencia, deben ser optimizados para cada situación en particular.

Desarrollo de una red neuronal artificial

El desarrollo adecuado de una red neuronal involucra una serie de consideraciones y de pasos que se discuten brevemente a continuación.

Entrenamiento

El entrenamiento de una red neuronal es el proceso a través del cual se ajustan los pesos, de manera que la respuesta de la misma se aproxime lo más posible a la respuesta esperada (conocida) por el investigador.

Ello quiere decir que, en esta primera etapa, es preciso suministrar tanto los datos de entrada (descriptores del problema) como los de salida (soluciones del problema). A cada pareja de datos entrada/salida se le conoce como “patrón”; en tanto que el proceso de entrenamiento ejecutado de la manera previamente mencionada, es del tipo “supervisado”.

Existe otro tipo de estrategia, la “no supervisada”, en la que sólo los datos de entrada son suministrados a la red, y se permite que ésta se establezca a través del empleo de algún tipo de algoritmo. Este último tipo de aprendizaje, sin embargo, no será discutido en el presente trabajo.

El proceso de ajuste de los pesos se lleva a cabo de manera iterativa, comparando las respuestas que arroja la red al suministrársele un grupo de datos de entrada con las respuestas proporcionadas a la misma, y buscando minimizar la diferencia entre ambas respuestas mediante algún algoritmo.

Una de las funciones empleadas para monitorear tal diferencia es el error cuadrático medio (ECM) (11):

$$ECM = \sum_i^n \frac{(Y_i^{Red} - Y_i^{Real})^2}{n} \quad (2)$$

en donde Y_i^{Red} y Y_i^{Real} , corresponden a las *i*-ésimas respuestas de la red y la suministrada por el usuario, respectivamente, y *n* es el número de patrones empleados en el entrenamiento. El ajuste de los pesos se ejecuta mediante diversas estrategias, una de las cuales se discutirá posteriormente.

Patrones: Características y transformaciones

La calidad del modelo generado por una red neuronal depende, en gran medida, de la calidad de los datos suministrados a la misma. Por esa razón, el investigador ha de hacer especial énfasis en la selección de los datos que se van a emplear en el entrenamiento de la misma. Ellos han de ser, primeramente, representativos del problema en cuestión. La escogencia equivocada de datos anómalos se va a traducir en un modelo neuronal errado. Es aconsejable, en segundo lugar, seleccionar una cantidad “grande” de patrones*, de manera que la red disponga de suficientes ejemplos con los cuales ajustar los pesos correspondientes.

Otro punto importante es el pre-tratamiento de los datos. En muchos casos se hace imprescindible transformar los datos de entrada; en particular cuando éstos varían, e. g., en las unidades de medición, magnitud, etc. Lo anterior se hace con el fin de evitar que la red le otorgue *a priori* una importancia indebida a ciertos datos basándose en aspectos que no son de relevancia para la solución del problema.

Las transformaciones pueden ser de varios tipos, a saber, normalización, estandarización, reducción mediante análisis por componentes principales, etc. Debido a que no existen reglas claras que permitan determinar cuál transformación puede ser la más adecuada para un problema dado, se hace imprescindible que el investigador evalúe a menudo varias de ellas antes de la selección final.

El pre-tratamiento de los datos de salida es menos complejo, por cuanto sólo puede haber dos casos: salidas binarias o continuas. En el primer caso los datos de salida son codificados mediante una combinación de 0 y/o ± 1 .

En el segundo caso, puede ser necesaria la normalización de los datos. Estos dos últimos aspectos serán posteriormente discutidos con más detalle (ver “Capacidades de las Redes Neuronales Artificiales”).

* Esta expresión se emplea exclusivamente para diferenciar a las unidades que conforman la capa de entrada de las unidades en las capas activas (neuronas), por cuanto las primeras, como ya se discutió: NO ejecutan ninguna función de transformación; SÓLO distribuyen los datos a las capas siguientes.

* El tamaño de la base de datos depende de la complejidad propia del sistema de estudio y de la disponibilidad de los mismos. Por esta razón, no existen reglas que le sugieran al usuario qué cantidad de datos mínimas serán necesarias para la resolución adecuada del problema.

La Regla Delta Generalizada

El ajuste de los pesos se lleva a cabo con el fin de minimizar la diferencia entre la respuesta de la red y la respuesta "real". Para esto es posible emplear distintas funciones de aprendizaje. Una de las más utilizadas para tal fin es, como ya se ha mencionado, el algoritmo de retro-propagación, también conocido como la *Regla Delta Generalizada*.⁶ Esta estrategia puede ser implementada en arquitecturas con conexión total o parcial entre neuronas de capas consecutivas, o entre capas no consecutivas (ver figura 4).

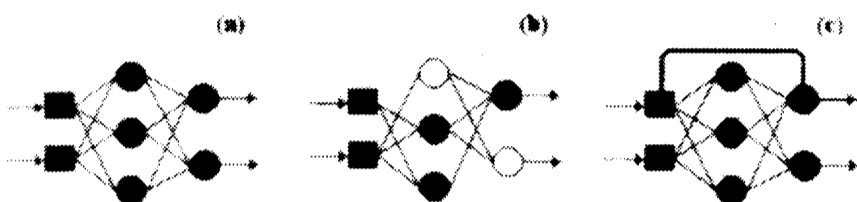


Figura 4. La regla delta generalizada puede ser implementada en arquitecturas con conexión: (a) total, (b) parcial (las neuronas en blanco no reciben información de todas las neuronas), o (c) entre capas no consecutivas (línea gruesa).

Debido al carácter general de este artículo vamos a dejar a un lado la derivación detallada de la misma, y nos conformaremos con presentar la forma general de esa expresión:³

$$\Delta w_{ji}^l = w_{ji}^{l(nuevo)} - w_{ji}^{l(anterior)} = \eta \delta_j^l Y_i^{(l-1)} + \mu [\Delta w_{ji}^{l(anterior)}] \quad (3)$$

en donde los sub-índices j e i representan a la neurona actual y a aquella de donde proviene la señal de entrada, respectivamente; el super-índice l indica la capa en la que se está haciendo el ajuste de los pesos; δ_j^l (cuya derivación, que no será discutida en este trabajo, depende de la capa en cuestión) es el error asociado a la j -ésima neurona; y Y_i es la respuesta de la i -ésima neurona. El término η , la *rata de aprendizaje*, es una variable que determina la velocidad de búsqueda en el espacio de soluciones (ver el párrafo siguiente). Por otra parte, el término a la extrema derecha de la igualdad, y que incluye al *factor de momento* (μ), es una modificación del algoritmo de retro-propagación que contribuye a "impulsar" la búsqueda en la dirección más favorable. Ambas variables (cuyos valores oscilan entre 0 y 1) deben ser determinadas empíricamente por el usuario de acuerdo a cada caso.

Algunos autores sugieren que la suma de los mismos debe ser menor o igual a uno,³ pero ello no necesariamente tiene que ser así.¹²

La minimización del error puede ilustrarse gráficamente (ver figura 5) como el ajuste de los pesos de manera de encontrar un mínimo (máximo) global en una superficie N-dimensional cuyo tamaño viene dado por el número de variables independientes involucradas en el problema (en el ejemplo hipotético de la Figura 5 se trata de un espacio bidimensional). Inicialmente se desconoce la dirección de búsqueda, por lo que es necesario asignar los valores numéricos de los pesos de forma aleatoria. En este sentido, es preciso destacar que el desempeño de una red neuronal puede depender de esta escogencia inicial (¡sobre la que no se tiene control!).⁷ Es por ello que se recomienda evaluar una misma arquitectura reiniciando los pesos repetidamente, y escoger finalmente aquella cuyo desempeño sea el más cercano al promedio. Una rata de aprendizaje muy pequeña se traducirá en una búsqueda más exhaustiva, pero también más lenta [figura 5(a)]. Por otra parte, un valor de ζ muy alto aumentará la velocidad de búsqueda, aunque a expensas de la minuciosidad con la que se ejecutará la misma. El factor de momento ha de ajustarse de forma que prevenga a la red de quedar atrapada en lo que se conoce como un *mínimo* (o *máximo*) *local* [figura 5(a)]. Un valor pequeño implica que se van a tomar menos en cuenta los cambios en los pesos que han ocurrido en los ciclos anteriores. Un valor mayor facilitará la salida de la red de los temidos mínimos (máximos) locales en la dirección adecuada, tal y como se representa en la figura 5(b).

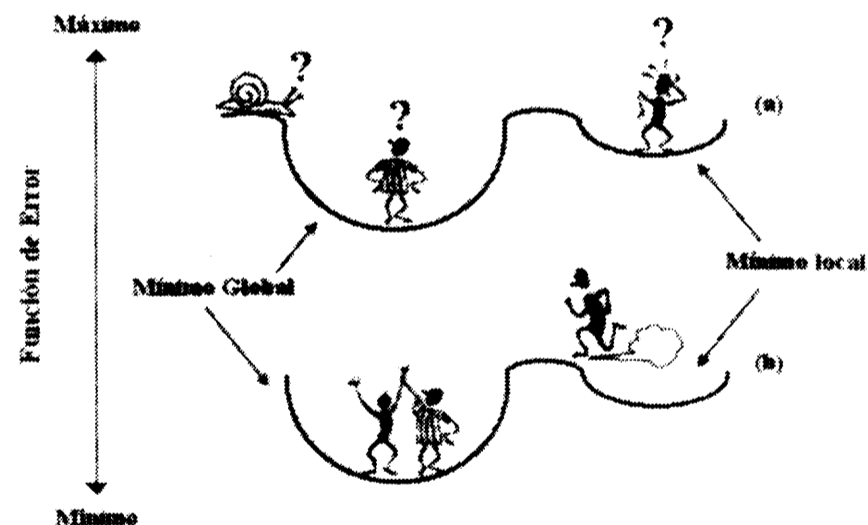


Figura 5. El problema de la búsqueda del mínimo global en el espacio de soluciones: (a) Sin la selección adecuada de la rata de aprendizaje y/o el factor de momento la red puede demorarse mucho en reducir el ECM, o puede quedar atorada en un mínimo local. (b) El uso de dichas variables optimizadas permite solventar dichos problemas.

El entrenamiento de la red es un proceso iterativo que concluye cuando se cumple una de las siguientes imposiciones del usuario: *i*) que los patrones de entrenamiento sean presentados repetidamente a la red por un número de ciclos (o *épocas*) dado; o, *ii*) que la función de error escogida alcance un valor (mínimo) pre-determinado. Sin embargo, y como se verá en las secciones que siguen a continuación, la red neuronal que resulte de esta etapa de entrenamiento no tiene que ser la más adecuada para

la tarea designada. Dicha red debe ser sometida a un proceso de validación y prueba previo a su implementación final.

Validación

Un punto importante que hay que tomar en consideración es la posibilidad de “sobre-entrenar” a la red. Esto se refiere al fenómeno que se observa cuando la red ajusta los pesos de las neuronas de forma de reproducir exactamente las respuestas suministradas por el usuario durante el proceso de entrenamiento. Cuando una red “memoriza” en lugar de “aprender”, pierde una de sus propiedades más atractivas: la capacidad de *generalización*. Ello se traduce en la imposibilidad de la misma para predecir respuestas plausibles ante un grupo de datos de entrada nuevos, i. e., que no han sido vistos por la red durante el entrenamiento. Para evitar el problema mencionado es conveniente utilizar un grupo de *datos de validación*, el cual ayuda a determinar el tiempo prudencial de entrenamiento.

La figura 6 sirve para ilustrar la importancia de esta etapa.¹² Se puede apreciar como el error cuadrático medio disminuye a medida que los datos de entrenamiento son procesados iterativamente por la red. Inicialmente hay una reducción brusca del error, entre 1000 y 2000 ciclos, lo cual puede estar relacionado con un cambio significativo en los pesos de las neuronas. Entre 2×10^3 y 10^4 ciclos continúa disminuyendo el ECM, si bien no con la misma rapidez. El usuario podría estar tentado a utilizar la red neuronal resultante tras 10^4 ciclos de entrenamiento, por cuanto el ECM es el menor. Sin embargo, de hacer esto, estaría incurriendo en un error que sólo podría ser apreciado a través de la inclusión de los datos de validación.

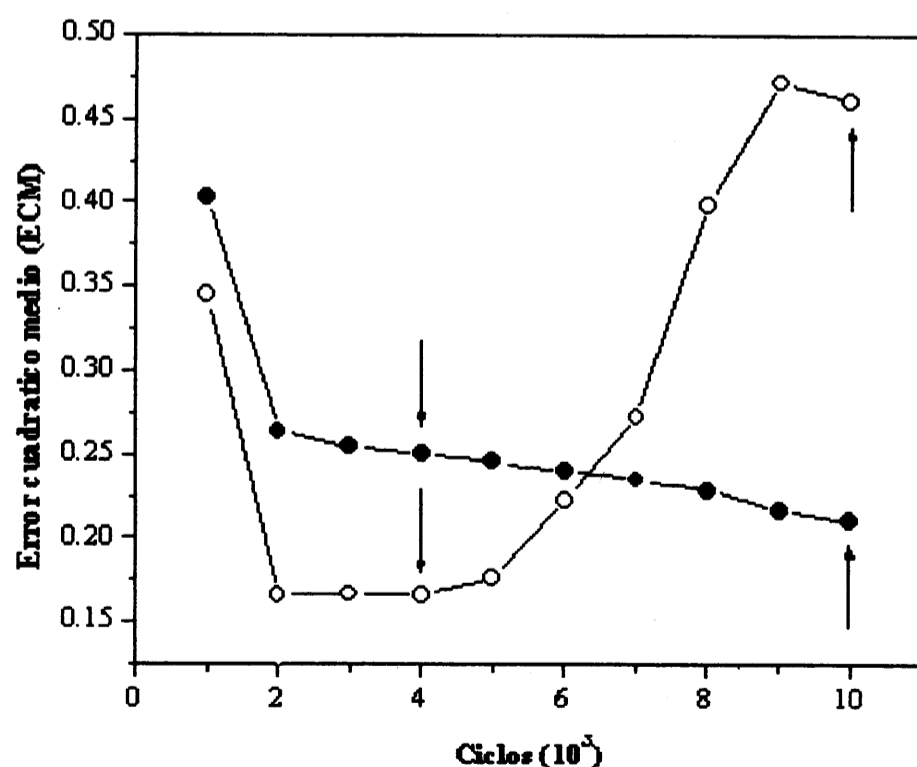


Figura 6. Variación del error cuadrático medio en función del número de iteraciones para un grupo de datos de entrenamiento (●) y uno de validación (○). Las flechas indican los puntos más importantes a considerar para la selección del tiempo de entrenamiento de la red neuronal (ver texto para más detalles).

En efecto, obsérvese como el ECM para este último grupo de datos, después de alcanzar un mínimo entre 2×10^3 y 4×10^3 , aumenta considerablemente hasta alcanzar un máximo entre 9×10^3 y 10^4 ciclos. Esto implica que la capacidad de generalización de la red se ha deteriorado significativamente a ese punto. Un buen compromiso es entonces detener el entrenamiento tras 4×10^3 ciclos, cuando se ha alcanzado el menor ECM para ambos grupos de datos.

Prueba

La última etapa en el desarrollo de una red neuronal consiste en evaluar su desempeño empleando datos que no han sido utilizados durante la etapa de entrenamiento. Los datos de prueba que se hayan seleccionado para este fin deben cumplir con un requisito fundamental, a saber, los mismos deben encontrarse dentro del rango de los datos empleados en la etapa previa. Ello se debe a que los modelos desarrollados por las redes neuronales, en analogía con aquellos desarrollados por medios matemáticos convencionales, permiten *interpolar* con un alto grado de exactitud. De igual manera, cuando se utilizan datos que se encuentran por fuera del rango de aquellos para los cuales fueron ajustados los pesos, es decir, se le pide a la red neuronal que *extrapole*, entonces se corre el riesgo de obtener resultados inexactos. Una analogía útil es esta: Una vez aprendidos, un niño es capaz de reconocer los caracteres que le permiten expresarse de manera escrita en su lengua nativa, e. g., el niño venezolano identificará bajo cualquier circunstancia las letras en español, a pesar de las variaciones que le imprimen diferentes personas al escribirlos. Pero ese conocimiento no le va a permitir reconocer el significado de un carácter chino: para ello es necesario “entrenarle” (enseñarle) dichos datos también.

Capacidades de las redes neuronales artificiales

La arquitectura de una red neuronal, así como su capacidad de adaptación, las convierten en herramientas valiosas en la resolución de problemas de muy diversas índoles. Seguidamente se describen algunas de las funciones que pueden desempeñar las redes neuronales multi-capas entrenadas con el algoritmo de retro-propagación.

Clasificación

La clasificación de un objeto dado consiste en determinar la pertenencia del mismo a un conjunto o categoría de objetos que comparten características similares, y que a su vez difiere de otros grupos. Para ello se pueden emplear algunos atributos de dichos objetos, o variables descriptivas, tales como las concentraciones de los elementos trazas, o el espectro de un espécimen dado, etc. Debido a que una red neuronal sólo puede procesar valores numéricos, resulta obligatorio codificar las categorías. Esto se puede llevar a cabo fácilmente mediante una combinación de 0,

$y/0 \pm 1$, dependiendo del número de neuronas y de la función de transferencia en la capa de salida.

La codificación de los datos de salida puede verse más claramente con un ejemplo. Supóngase que un analista desea determinar cuál de entre tres casas comerciales (A, B y C) fabrica un medicamento "X", mediante redes neuronales radiales, empleando como dato de entrada el cromatograma de dicho compuesto. La salida de la red será entonces la casa comercial. En este caso se puede utilizar una capa de salida con tres neuronas. El código que identifique a cada fabricante, de emplearse una función de transferencia sigmoideal, podría ser A (1,0,0), B(0,1,0) y C(0,0,1). Durante el entrenamiento, la red neuronal buscará entonces ajustar los pesos de forma que la señal de salida, por ejemplo en el caso de una función de transferencia sigmoideal, se encuentre en la región de mínima ($Y_i = 0$) o máxima ($Y_i = 1$) respuesta (ver figura 2).

Algunas de las aplicaciones de las redes neuronales en la resolución de este tipo de problema lo constituyen la determinación de la procedencia de bebidas alcohólicas fabricadas ilegalmente (Aguardiente de Cocuy) a partir de las concentraciones de hierro, cobre y cinc;¹² en la determinación del origen geográfico de una *muestra de especímenes** de aceite de oliva extra virgen proveniente de distintas regiones de Italia, a través de su análisis por espectrometría de masas;¹³ como herramienta de control de calidad de un producto farmacéutico contenido de l-triptofano (para el tratamiento de la obesidad, el insomnio y la depresión), y que es producido por seis fabricantes diferentes;¹⁴ y en la caracterización geoquímica de una muestra de sedimentos mediante análisis por espectroscopía de reflectancia difusa.¹⁵

Representación gráfica

Como su nombre lo sugiere, ésta consiste en la proyección en dos o tres dimensiones, de grupos de objetos dados que pueden compartir o no algunas características que los describen. Lo anterior es particularmente útil cuando se estudia el efecto de más de tres (3) variables sobre la respuesta de un sistema. La estructura original y la información más relevante de los datos no son alteradas, a pesar de dicha transformación, gracias a lo cual es posible establecer relaciones que no se pueden apreciar de otra manera (¡Trate de resolver un problema en cuatro dimensiones solamente y comprenderá la importancia de la reducción espacial!).

* Desde el punto de vista estadístico, una *muestra* se refiere a un grupo (de tamaño finito) de objetos tomados de una población dada, e. g., una muestra de balones aforados de una fábrica de material de laboratorio. Cada objeto de dicha muestra es, por el contrario, un *especimen*. Ambos términos, sin embargo, suelen emplearse comúnmente sin distinción.

La red de tipo "cuello de botella", cuya topología está esquematizada en la figura 3, puede ser empleada para la representación gráfica en dos o tres dimensiones, de acuerdo con el número de neuronas que se empleen en la segunda capa escondida.^{16,17} Este caso, así como en los que se discuten seguidamente, la red neuronal procurará ajustar los pesos para que la señal resultante se encuentre en el intervalo (continuo) de transición mostrado en la figura 2. La identificación en dos dimensiones de una muestra de especímenes de leche de vaca contaminadas con dioxinas provenientes de cinco regiones de Holanda, se encuentran entre las aplicaciones en este campo.¹⁶

Modelado

Las redes neuronales son capaces de modelar sistemas, lineales y/o no lineales, mediante el ajuste de los pesos de las neuronas, y con el empleo de las funciones de transferencia adecuadas. Ello representa una importante ventaja frente a los modelos matemáticos convencionales, particularmente en aquellos casos en los que no se conoce(n) *a priori* la(s) posible(s) relación(es) entre las variables dependiente(s) e independiente(s), o la complejidad del problema es tal que dificulta, e incluso imposibilita, la formulación de dicha(s) relación(es).

Se pueden citar como ejemplos el modelado de una curva de calibración no lineal para la determinación de cadmio por GFAAS,¹⁸ el comportamiento de retención de diversos analitos en mezclas de metanol- tetrahidrofurano-agua o metanol-acetonitrilo-agua para su determinación por HPLC,¹⁹ y el modelado de las respuestas electroquímicas de la adenina y de la citosina en la región de evolución de hidrógeno para su determinación simultánea mediante LSV y DPP.²⁰

Optimización

Este aspecto está también relacionado con el discutido previamente. El problema de optimizar consiste en encontrar aquellas condiciones en las que un grupo de variables independientes conducen a la obtención de un máximo o de un mínimo en una variable dependiente, o variable respuesta. De esto se deduce que, previo a la optimización de un sistema, es necesario disponer del modelo que describe la relación entre las variables involucradas.

Este modelo, como se vio previamente, puede ser extraído de los datos por la misma red neuronal.

Entre las aplicaciones se pueden nombrar la optimización de las condiciones analíticas y/o instrumentales para la determinación de aluminio²¹ y rutenio,²² empleando sistemas de inyección en flujo y detección espectrofotométrica; la determinación de impurezas en drogas mediante cromatografía de gases;²³ y la selección de la fase móvil en cromatografía de par iónico.²⁴

Problemas asociados a las redes neuronales artificiales

Como toda herramienta, las redes neuronales artificiales no están exentas de problemas. Una de las críticas más importantes deriva de su carácter de "cajas negras". En algunos casos, particularmente en aquellos en los que el número de neuronas es considerable (desde varias decenas hasta cientos de miles) no es posible "seguir el rastro" de los datos introducidos a través de la capa de entrada. Hay que recordar en este sentido que cada uno de los datos de entrada es transferido a TODAS las neuronas de la capa siguiente, y posteriormente transformado como se explicó en una sección previa. Por esta misma razón, resulta imposible a su vez determinar la importancia de los pesos neuronales, de manera análoga a como se podría hacer, por ejemplo, con los coeficientes de una función lineal discriminante.¹² Adicionalmente, los mismos pesos no poseen un equivalente físico, tal y como lo pueden tener, por ejemplo, el intercepto y la pendiente de una regresión lineal empleada para cuantificar en espectroscopía de absorción atómica.¹⁸ Todo lo anterior se traduce en la incapacidad del usuario de establecer lo que ocurre, matemáticamente hablando, dentro de una red neuronal artificial.

Por otra parte, es preciso destacar el hecho que, para que la red pueda establecer adecuadamente las relaciones correspondientes entre los datos de entrada y salida, es necesario tener un volumen suficientemente grande de datos que permita su segmentación en grupos de entrenamiento, validación y prueba (¡el tamaño SI importa!). Cabe resaltar que no existen reglas pre-establecidas que permitan determinar cuán grande ha de ser la base de datos a emplear en la resolución de un problema en particular. En cualquier caso, su adquisición puede ser una tarea complicada en sí misma, dependiendo de aspectos tales como tiempo y costos de análisis, etc.

El uso de la regla delta generalizada como algoritmo de aprendizaje trae consigo dos problemas. Por una parte es posible que el empleo de la misma no permita la consecución del mínimo/máximo global de la función en el espacio de soluciones N-dimensional o, por el contrario, que la red se "atasque" en un mínimo/máximo local (figura 5). Por otra parte, se ha verificado que la regla delta generalizada prolonga el tiempo de desarrollo (entrenamiento) de las redes neuronales artificiales.²⁵

Nada de lo anterior, sin embargo, ha detenido a los químicos analíticos en su búsqueda de nuevas aplicaciones. Como lo han planteado algunos autores,¹⁸ la veracidad del modelo (en términos físicos) desarrollado por la red puede no ser tan importante como el hecho que la misma pueda ejecutar adecuadamente la tarea para la cual ha sido diseñada. Por otra parte, la disponibilidad de técnicas instrumentales cada vez más sofisticadas usualmente ha venido acompañada de un volumen cada vez mayor de datos derivados de su uso, por lo que la obtención de datos no parece

representar un problema *per se*. Finalmente, el hecho que la red neuronal pueda desempeñar su función de forma casi inmediata, una vez ha sido entrenada, compensa hasta cierto punto el problema asociado al tiempo que consume el desarrollo de la misma.

Conclusiones

Las redes neuronales artificiales pertenecen al grupo de técnicas de computación inteligente que pueden ser empleadas en la resolución de numerosos tareas, e. g., clasificación, representación gráfica, modelado y optimización. Una de las ventajas más importantes de estas herramientas es que su implementación no requiere del conocimiento *a priori* del sistema a estudiar. Lo anterior, sin embargo, no quiere decir que las mismas sean una suerte de panacea para la resolución de todos los problemas. La aplicación y el correcto desempeño de las mismas dependerán, como ocurre en cualquier caso, de las necesidades del usuario y del escrutinio con el cual éste lleve a cabo su desarrollo.

Agradecimientos

Uno de los autores (E.A.H.C.) agradece al Fondo Nacional de Ciencia, Tecnología e Innovación (FONACIT) por la beca otorgada para llevar a cabo estudios de Doctorado en Química Analítica en la Universidad de Los Andes (Mérida, Venezuela).

Referencias

1. S. D. Brown, R. S. Bear, Jr., Chemometric techniques in electrochemistry: a critical review, *Crit. Rev. Anal. Chem.* **24**, 99-131, 1993.
2. J. Aguilar Castro, F. Rivas Echeverría (Eds.), Introducción a las Técnicas de Computación Inteligente, Universidad de Los Andes, Venezuela, 2001.
3. J. Zupan, J. Gasteiger, Neural Networks in Chemistry and Drug Design, Wiley-VCH, 2nd Ed., Germany, 1999.
4. N. K. Bose, P. Liang, Neural Networks Fundamentals with Graphs, Algorithms, and Applications, McGraw-Hill, Inc., USA, 1996.
5. S. Haykin, Neural Networks. A Comprehensive Foundation, Macmillan College Publishing Company, Inc., USA, 1994.
6. B. J. Wythoff, Backpropagation neural networks. A tutorial, *Chemom. Intell. Lab. Syst.* **18**, 115-155, 1993.
7. J. R. M. Smith, W. J. Melssen, L. M. C. Buydens, G. Kateman, Using artificial neural networks for solving chemical problems. Part I. Multi-layer feed-forward networks, *Chemom. Intell. Lab. Syst.* **22**, 165-189, 1994.
8. J. A. Burns, G. M. Whitesides, Feed-forward neural networks in chemistry: mathematical systems for classification and pattern recognition, *Chem. Rev.* **93**, 2583-2601, 1993.

9. J. Gasteiger, J. Zupan, Neural networks in chemistry, *Angew. Chem. Int. Ed. Engl.* **32**, 503-527, 1993.
10. J. Zupan, J. Gasteiger, Neural networks: a new method for solving chemical problems or just a passing phase?, *Anal. Chim. Acta*, **248**, 1-30, 1991.
11. Propagator™, Neural Network Development Software, Version 1.0 for Windows, ARD Corp., USA, 1993.
12. E. A. Hernández-Caraballo, R. M. Avila-Gómez, T. Capote, F. Rivas, A. G. Pérez, Classification of Venezuelan spirituous beverages by discriminant analysis and artificial neural networks based on their Zn, Cu and Fe concentrations, *Talanta*, **60**, 1259-1267, 2003.
13. G. J. Salter, M. Lazzari, L. Giansante, R. Goodacre, A. Jones, G. Surrichio, D. B. Kell, G. Bianchi, Determination of the geographical origin of Italian extra virgin olive oil using pyrolysis mass spectrometry and artificial neural networks, *J. Anal. Appl. Pyrolysis*, **40-41**, 159-170, 1997.
14. W. J. Welsh, W. Lin, S. H. Tersigni, E. Collantes, R. Duta, M. S. Carey, W. L. Zielinski, J. Brower, J. A. Spencer, T. P. Layloff, Pharmaceutical fingerprinting: Evaluation of neural networks and chemometric techniques for distinguishing among same-product manufacturers, *Anal. Chem.*, **68**, 3473-3482, 1996.
15. T. Udelhoven, B. Schütt, Capability of feed-forward neural networks for a chemical evaluation of sediments with diffuse reflectance spectroscopy, *Chemom. Intell. Lab. Syst.*, **51**, 9-22, 2000.
16. P. Cleij, P. Hoogerbrugge, Linear data projection using a feedforward neural network, *Anal. Chim. Acta*, **348**, 495-501, 1997.
17. M. Daszykowski, B. Walczak, D. L. Massart, Projection methods in chemistry, *Chemom. Intell. Lab. Syst.*, **65**,: 97-112, 2003.
18. E. A. Hernández-Caraballo, R. M. Avila-Gómez, F. Rivas, M. Burguera, J. L. Burguera, Increasing the working calibration range by means of artificial neural networks for the determination of cadmium by graphite furnace atomic absorption spectrometry, *manuscrito en revision (Talanta 2003)*.
19. R. H. Zhao, B. F. Yue, J. Y. Ni, H. F. Zhou, Y. K. Zhang, Application of an artificial neural network in chromatography- retention behavior prediction and pattern recognition, *Chemom. Intell. Lab. Syst.*, **45**, 163-170, 1999.
20. E. Cukrowska, L. Trnková, R. Kizek, J. Havel, Use of artificial neural networks for the evaluation of electrochemical signals of adenine and cytosine in mixtures interfered with hydrogen evolution, *J. Electroanal. Chem.*, **503**, 117-124, 2001.
21. Y. Zhou, H. Wang, G. Sun, Y. Fan, X. Chen, Z. Hu, Application of artificial neural networks in multifactor optimization of an FIA system for the determination of aluminum, *Fresenius J. Anal. Chem.*, **366**, 17-21, 2000.
22. H. Wang, Y. Zhou, Y. Zhao, Q. Li, X. Chen, Z. Hu, Optimization of on-line microwave flow injection analysis system by artificial neural networks for the determination of ruthenium, *Anal. Chim. Acta*, **429**, 207-213, 2001.
23. W. Guo, P. Zhu, H. Brodowsky, The study for optimization of chromatographic condition by means of artificial neural networks, *Talanta*, **44**, 1995-2001, 1997.
24. J. V. S. Gobburu, W. L. Shelver, W. H. Shelver, Application of artificial neural networks in the optimization of HPLC mobile-phase parameters, *J. Liquid Chrom.*, **18**, 1957-1972, 1995.
25. E. P. P. A. Derks, L. M. C. Buydens, Aspects of training and validation on noisy data. Part 1. Training aspects, *Chemom. Intell. Lab. Syst.*, **41**, 171-184, 1998.

Edwin A. Hernández-Caraballo*, **Rita M. Avila-Gómez**** y **Franklin Rivas*****

*Instituto Venezolano-Andino de Investigaciones Químicas,
Faculta de Ciencias,
Universidad de Los Andes, Mérida 5101-A, Venezuela
**Departamento de Procesos Agroindustriales,
Decanato de Agronomía,
Universidad Centro-occidental Lisandro Alvarado,
Núcleo Obelisco, Barquisimeto 3002, Venezuela
***Laboratorio de Sistemas Inteligentes,
Escuela de Ingeniería de Sistemas, Facultad de Ingeniería,
Universidad de Los Andes, Mérida 5101, Venezuela

TESIS

Licenciatura en Química
2003

Título: ESTUDIO DE LA REACCIÓN DE ESTERIFICACIÓN DEL ÁCIDO P-HIDROXIBENZOICO CON METANOL Y BUTANOL SOBRE CATALIZADORES ÁCIDOS SÓLIDOS

Autor: Crespo P. Isis Garibay **Tutor:** Mireya De Goldwasser

Institución: Escuela de Química, Facultad de Ciencias, Universidad Central de Venezuela

Título: EVALUACIÓN DE UN PROTOTIPO DE CÁMARA PARA LA INTRODUCCIÓN DIRECTA DEL AEROSOL GENERADO POR UN MICRONEBULIZADOR

Autor: Da Corte Barreiro, John H **Tutor:** José Chirinos

Institución: Escuela de Química, Facultad de Ciencias, Universidad Central de Venezuela