

EL ROL DEL ALGORITMO DE ENTRENAMIENTO EN LA SELECCIÓN DE MODELOS DE REDES NEURONALES

ROLE OF THE TRAINING ALGORITHM IN MODEL SELECTION ON NEURAL NETWORKS

Paola Sánchez¹, Juan Velásquez²

¹Ph.D. (c). Escuela de Sistemas. Facultad de Minas. Universidad Nacional de Colombia. Carrera 80 No. 65 – 223. Medellín, Colombia. Autor para Correspondencia. E-mail: pasanche@unal.edu.co ²Ph.D. Escuela de Sistemas. Facultad de Minas. Universidad Nacional de Colombia. E-mail: jvelasq@unal.edu.co.

Rev. U.D.C.A Act. & Div. Cient. 14(1): 149 - 156, 2011

RESUMEN

La capacidad de ajuste de una red neuronal se ve a menudo afectada por la configuración usada, en especial, en relación al número de neuronas ocultas y de variables de entrada, toda vez que, a medida que el número de parámetros del modelo aumenta, se favorece el aprendizaje de la red y, por tanto, el ajuste es mejor. Teóricamente, un proceso constructivo de adición de parámetros debería conducir a reducciones sistemáticas en el error, siempre y cuando, los modelos sean anidados en cada paso del proceso. En este trabajo, se valida la hipótesis que la adición de neuronas ocultas en modelos anidados debe conducir a reducciones en el error, sin importar el algoritmo de entrenamiento usado; para ejemplificar la discusión, se usaron la serie de pasajeros en líneas aéreas y de manchas solares de Box & Jenkins y los métodos de entrenamiento de Regla Delta y RProp. La evidencia experimental demuestra que los métodos de entrenamiento evaluados exhiben comportamientos diferentes a los teóricamente esperados, incumpliendo el supuesto de reducción del error.

Palabras clave: Redes Neuronales, Algoritmo de Entrenamiento.

SUMMARY

The Neural net's fit ability is often affected by the network configuration, particularly the number of hidden neurons and input variables. As the size of these parameters increases, the learning also increases, then the fit of network is better. Theoretically, if parameters are increasing regularly, the error should be reduced systematically, provided that the models

are nested for each step of the process. In this work, we validated the hypothesis that the addition of hidden neurons in nested models lead to systematic reductions in error, regardless of the learning algorithm used; to illustrate the discussion we used the number of airline passengers and Sunspots in Box & Jenkins, and RProp and Delta Rule as learning methods. Experimental evidence shows that the evaluated training methods show different behaviors as those theoretically expected, it means, not fulfilling the assumption of error reduction.

Key words: Artificial neural networks, training algorithm.

INTRODUCCIÓN

El pronóstico de series de tiempo es un área que ha despertado el interés de investigadores en diversos campos del conocimiento. Esta motivación ha conducido al surgimiento de una amplia gama de modelos, donde las redes neuronales artificiales, vistas como una técnica no paramétrica y no lineal de regresión, han demostrado un enorme potencial, debido a sus capacidades de adaptabilidad, generalización, aprendizaje y capacidad para representar relaciones no lineales (Zhang *et al.* 1998; Velásquez *et al.* 2008). El creciente interés en el desarrollo de aplicaciones de pronóstico con redes neuronales, se denota con la publicación de más de 5000 artículos de investigación presentes en la literatura (Crone & Kourentzes, 2009; Adya & Collopy, 1998).

Si bien, existen muchos reportes exitosos sobre el uso de las redes neuronales en el pronóstico de series de tiempo, también y tal como lo enuncian Zhang *et al.* (1998), se han reportado resultados que contradicen estas evidencias, generando así

inconsistencias sobre este tópico. Una explicación aceptada de las razones que conducen a resultados inconsistentes está relacionada con el elevado número de factores determinantes en el proceso de construcción del modelo. Dicho proceso de construcción incluye:

- La selección del algoritmo de aprendizaje o entrenamiento utilizado y de los valores de sus parámetros, los cuales, impactan, directamente, aspectos, como el tiempo de convergencia, el valor de los estadísticos de ajuste y la robustez del modelo final (Hagan *et al.* 1996; Fahlman, 1989; Tang & Koehler, 1994).
- La selección del número adecuado de neuronas en la capa oculta (Zhang *et al.* 2001; Murata *et al.* 1994).

Se sabe que estos aspectos están interrelacionados y que influyen en la bondad del modelo para ajustarse a los datos, así como en su capacidad de generalización. Qi & Zhang (2001) analizan la influencia del número de neuronas ocultas en el ajuste del modelo, pero no presentan resultados concluyentes; por lo que no existe, una visión clara, coherente y comúnmente aceptada sobre esta interrelación. La multiplicidad de criterios en cada una de las fases del proceso de construcción del modelo hace que la técnica sea inestable, ya que cualquier cambio en un solo criterio conduce a grandes cambios en el ajuste. Consecuentemente, un análisis del comportamiento de estos factores es de especial interés a la hora de implementar un modelo de red neuronal.

Se sabe que un proceso adecuado para la estimación de los parámetros de una red neuronal es el punto de partida para determinar la forma del modelo, toda vez que se parte del supuesto que dicha estimación es óptima. Con respecto a esto, se ha demostrado que las redes neuronales artificiales son aproximadores universales de funciones, tal como lo establecen Hornik *et al.* (1989), basado en el teorema de superposición de Kolmogorov, un perceptrón multicapa con un número suficiente de neuronas ocultas puede aproximar, con un nivel arbitrario de precisión, cualquier función continua definida en un espacio finito; sin embargo, la demostración matemática no da indicios de cómo se deben estimar los parámetros de la red neuronal tal que el error se pueda llevar arbitrariamente a cero; igualmente, tampoco presenta indicios que indiquen si el error de ajuste para un modelo más complejo (con más neuronas de entrada o más neuronas ocultas), respecto a un modelo más simple, debería aumentar, disminuir o permanecer igual. Uno de los intereses de este artículo es mostrar que el error de ajuste a la muestra de entrenamiento debe disminuir o, a lo sumo, permanecer igual, cuando se aumentan las neuronas de la capa oculta o de la capa de entrada; no obstante, este aspecto no es discutido profundamente en la literatura.

El primer objetivo de este artículo fue analizar las implicaciones teóricas y conceptuales de esta propiedad y los requerimientos que de ella se derivan, los cuales, deberían ser cumplidos por las diferentes metodologías de estimación de modelos de redes neuronales artificiales.

Se debe anotar, que pareciera que la comunidad científica diera por cierto que siempre es posible obtener un modelo que cumpla con el grado de precisión deseado en el ajuste, de tal forma que, el primer objetivo planteado, careciera de importancia real. Así, la contribución realizada en esta investigación consiste en analizar este aspecto, para determinar sus implicaciones sobre el proceso de obtención de un modelo de redes neuronales, con el fin de demostrar que se presentan consideraciones conceptuales con profundas implicaciones prácticas, que no han sido contempladas.

Por otra parte, se sabe que una de las razones del desempeño pobre de una red neuronal está relacionada con las fortalezas y las falencias del algoritmo de entrenamiento utilizado para calcular sus pesos o parámetros; esto es, su capacidad para eludir los mínimos locales de la función de error de ajuste y la facilidad de calibración de sus parámetros, para aplicarlo a cada caso particular. Este aspecto ha motivado el planteamiento de muchos algoritmos novedosos, para la optimización que realizan una búsqueda computacionalmente más rápida del óptimo o que permitan encontrar óptimos de mejor calidad (mejor ajuste del modelo), que otros métodos disponibles; no obstante, en el desarrollo de dichos algoritmos no se tienen en cuenta las implicaciones prácticas que las redes sean aproximadores universales de funciones y de la reducción del error de ajuste, a medida que se aumentan neuronas en la capa oculta.

El segundo objetivo de este trabajo fue analizar, de forma empírica, si dos de los principales algoritmos de optimización usados para entrenar redes neuronales —regla delta generalizada y *resilient back propagation* (Rprop)— cumplen con las implicaciones derivadas en el primer objetivo, en términos de la reducción del error de ajuste. Existen dos aportes fundamentales en la consecución de este objetivo: primero, el tipo de análisis planteado nunca ha sido realizado en la literatura técnica, de tal forma que, muchas razones que se podrían dar, en principio, para rechazar este objetivo son basadas en la experiencia práctica más que en un proceso riguroso de experimentación, por lo que este rechazo carecería de fundamentos sólidos. Y segundo, se presenta una nueva perspectiva del problema avalada por resultados experimentales que corroboran el análisis efectuado.

El análisis derivado de las conclusiones que se pueden lograr de los dos objetivos anteriores tiene profundas e importantes

implicaciones conceptuales, teóricas y prácticas del proceso de selección del modelo, esto es, sobre las diferentes metodologías que se han desarrollado para determinar el número adecuado de neuronas en la capa oculta. El tercer objetivo de esta investigación es realizar dicho análisis.

La originalidad, importancia y relevancia de este trabajo está basada en los siguientes aspectos:

1. Se evalúan aspectos teóricos de la construcción de modelos de redes neuronales, tales como el cumplimiento del supuesto de reducción del error en un proceso de adicción de neuronas ocultas y la inferencia del método de entrenamiento en los resultados.
2. Se establecen criterios que deben ser considerados a la hora de construir una representación de series de tiempo con redes neuronales, que son obviados en su implementación y se discute cómo estos afectan el modelo final obtenido.

MATERIALES Y MÉTODOS

El perceptrón multicapa: En la figura 1, se presenta el esquema representativo de una red perceptrón multicapa. Este tipo de arquitectura de red neuronal consta de:

- Una capa de entrada, donde existe una neurona por cada valor rezagado de la serie de tiempo y_t , con un número de P retardos.
- Una o más capas ocultas. En el caso particular analizado, se consideró únicamente una capa oculta con H neuronas. Cada neurona de la capa oculta transforma su entrada neta usando una función no lineal $g(\cdot)$, conocida técnicamente como función de activación o de transferencia, y la envía hacia la siguiente capa, a través

de las conexiones β_h .

- Una capa de salida. Para este caso, la capa de salida consta de una sola neurona, que corresponde al valor actual de la serie de tiempo y_t .

Tanto las capas ocultas como la capa de salida reciben un pulso unitario de una neurona que no recibe entradas, la cual, es notada en la figura 1, con la letra B . Las conexiones por las que se transmite este pulso son notadas como ω_h , para la capa oculta y, como η , para la neurona de la capa de salida.

Así, el valor actual de una serie de tiempo es una función no lineal de sus valores pasados y_{t-1}, \dots, y_{t-p} , la cual, se define como:

$$y_t = \eta + \sum_{h=1}^H \left\{ \beta_h \times g \left[\omega_h + \sum_{p=1}^p \alpha_{p,h} y_{t-p} \right] \right\} + e_t$$

donde e_t representa los errores o residuos del modelo que son independientes e idénticamente distribuidos (iid) con media cero y varianza constante σ_e^2 .

Con el diseño de una red neuronal artificial, se pretende conseguir que, para ciertos valores rezagados de la variable explicada, ésta sea capaz de aproximar el valor actual de la serie de tiempo con una precisión deseada (Zhang *et al.* 1998). Para ello, además de una estructura adecuada (determinada por los valores escogidos de P y H), se requiere de un proceso de aprendizaje, que permita modificar los valores de los pesos asociados a las distintas conexiones [$\eta, \omega_h, \beta_h, \alpha_{p,h}$ para $h = 1, \dots, H$ y $p = 1, \dots, P$], también conocidos como parámetros del modelo. Tal como lo

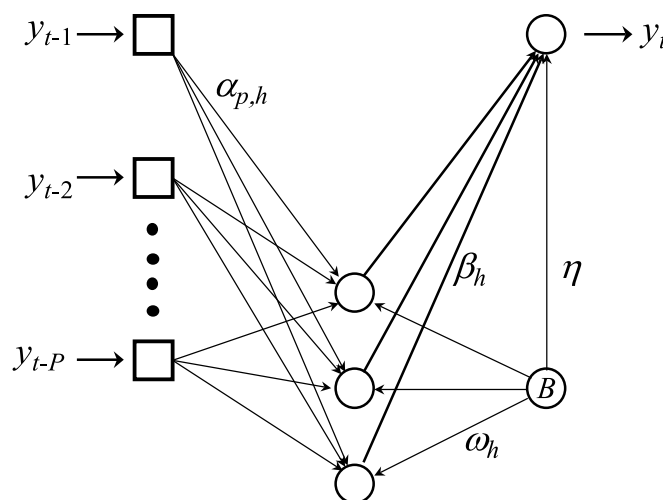


Figura 1. Esquema de un perceptrón multicapa con una capa oculta de procesamiento.

exponen Qi & Zhang (2001) existe una estrecha relación entre el desempeño del modelo y la selección de los valores de P y H . La importancia primaria de hacer una selección adecuada radica en las dificultades de convergencia del algoritmo de aprendizaje, que puede acarrear el incluir retardos irrelevantes y obtener un modelo final con pobre desempeño en ajuste y en generalización.

La literatura muestra el desarrollo de métodos constructivos, que permiten la selección del número de neuronas ocultas dentro del proceso de entrenamiento, mediante una evaluación de la conveniencia de adicionar o no un nuevo parámetro a la red, según éste disminuya el término del error. El desarrollo bajo el esquema constructivo exige que el error disminuya a medida que se adicionan parámetros al modelo; sin embargo, la base conceptual y las implicaciones prácticas de dicha reducción no son tenidas en cuenta a menudo.

Complejidad del modelo y reducción del error de ajuste:

Existen varios trabajos teóricos que demuestran que una red neuronal, tipo perceptrón multicapa, es un aproximador universal de funciones; un ejemplo es el trabajo de Hornik *et al.* (1989). En esta sección, se muestra que el error de ajuste debe disminuir o, en el peor caso, permanecer igual, cuando se adicionan neuronas ocultas o de entrada. El razonamiento es el siguiente:

- Sea un perceptrón multicapa con H neuronas en la capa oculta y que usa como entradas los primeros P rezagos de la serie. Se aplica la notación MLP (P, H), para representar este modelo.
- Los parámetros del modelo MLP (P, H) fueron estimados minimizando alguna función de error medida sobre un conjunto fijo de patrones de entrenamiento. Este error es notado como $E(P, H)$.
- Sea un perceptrón multicapa con una neurona oculta adicional; esto es, MLP ($P, H + 1$). Se cumple que $E(P, H) \geq E(P, H + 1)$, lo cual, se puede demostrar usando una contradicción. Supóngase que $E(P, H) < E(P, H + 1)$, en teoría esta situación no se puede dar. El modelo MLP ($P, H + 1$), se puede obtener del modelo MLP (P, H), al agregar una neurona oculta adicional; si se conservan los valores de los pesos del modelo MLP (P, H) y se hace cero la conexión β_{H+1} (independientemente de los pesos de las conexiones de la capa de entrada a la capa oculta), se tiene que $E(P, H) = E(P, H + 1)$, sin aplicar ningún proceso de optimización para ajustar los parámetros del modelo con una neurona adicional. Consecuentemente, el error debe permanecer igual o reducirse al agregar neuronas en la capa oculta.
- Sea un perceptrón multicapa con una entrada adicional; esto es, MLP ($P + 1, H$). Se cumple que $E(P, H) \geq E(P + 1, H)$. Al igual que en el caso anterior, el modelo MLP

($P + 1, H$), se puede obtener del modelo MLP (P, H) al agregar una nueva neurona de entrada para el rezago $P + 1$; si se conservan los valores de los pesos del modelo MLP (P, H) y se hacen cero las conexiones $\alpha_{P+1, h}$ para $h = 1, \dots, H$, se tiene que $E(P + 1, H) = E(P, H)$, sin aplicar ningún proceso de optimización para ajustar los parámetros del modelo con una entrada adicional. Consecuentemente, el error debe permanecer igual o reducirse al agregar nuevas entradas al modelo.

Lo anterior implica que un proceso de adición de neuronas ocultas va permitir (al menos teóricamente) una reducción secuencial del error de ajuste del modelo hasta un nivel arbitrariamente cercano a cero; en otras palabras, un modelo con más neuronas ocultas se debería ajustar mejor a los datos de entrenamiento que un modelo con menos neuronas ocultas; sin embargo, usualmente, los algoritmos de optimización no tienen en cuenta, dentro de su especificación, que se garantice dicha mejora, sugiriendo que la optimización, para cada posible modelo, se realiza de forma independiente.

En la literatura estadística, la reducción del error de ajuste al aumentar la complejidad del modelo es un concepto muy conocido; Hamilton (1994) es un ejemplo. El modelo MLP (P, H) es llamado modelo restringido, sub-modelo o modelo anidado respecto a los modelos MLP ($P, H + 1$) y MLP ($P + 1, H$), que son conocidos como completos. De aquí en adelante, el modelo completo, se refiere a MLP ($P, H + 1$), a menos que se especifique lo contrario. La comparación entre el modelo completo y el restringido es usada en el contraste del radio de verosimilitud, para determinar si una neurona oculta adicional es estadísticamente significativa —una explicación extensa es presentada por Hamilton (1994)—.

En conclusión, para un perceptrón multicapa y bajo un esquema constructivo de adición de neuronas ocultas, se debería presentar una reducción sistemática del error de ajuste, cada vez que se agrega una neurona en la capa oculta, hasta un nivel de precisión del ajuste arbitrariamente cercano a cero.

Información utilizada: Con el fin de validar empíricamente la reducción de del error de ajuste, se desarrollaron experimentos con dos diferentes conjuntos:

1. Uso de la serie del número mensual de pasajeros en líneas aéreas (Airline), la cual, consta de 144 observaciones que equivalen a los datos mensuales, entre 1949:01 y 1960:12. Para la serie Airline, se siguieron los lineamientos señalados por Faraway & Chatfield (1998) y Ghiassi *et al.* (2005), transformando la serie, mediante la función logaritmo natural, antes de llevar a cabo la experimentación.

2. Uso de la serie anual de manchas solares (SunSpots), que corresponde a 221 datos, en el periodo 1700 y 1920. Los datos de la serie Sunspots no fueron transformados.

Experimento 1: Se pretende evaluar si el ajuste de la serie es el mismo sin depender del algoritmo de entrenamiento que se use. Para ambas series, se consideraron configuraciones de red neuronal fijas, las cuales, se especifican por sus respectivos valores de P y H , y se estimaron los valores óptimos de los parámetros empleando los algoritmos de entrenamiento regla delta generalizada y Rprop. El método de la regla delta fue elegido por ser el algoritmo más comúnmente usado en las aplicaciones descritas en la literatura técnica para el entrenamiento de redes neuronales, mientras que el algoritmo RProp fue seleccionado por ser considerado uno de los algoritmos con más alto nivel de desempeño, velocidad de convergencia y robustez (Anastasiadis *et al.* 2003).

Para evitar la dependencia de los puntos iniciales y garantizar aleatoriedad en los resultados, cada proceso fue repetido 50 veces y se conservaron los valores de los pesos, que presentaron el menor error de ajuste.

Experimento 2: El objetivo fue evaluar si a medida que se adicionan neuronas en la capa oculta y manteniendo un número de entradas fijas, se produce una reducción del error de ajuste de los modelos a la serie de tiempo.

El experimento fue conducido de la siguiente forma:

- Para la serie Airline, se consideraron, como entradas, los rezagos 1, 12 y 13 (parámetros usados por Faraway & Chatfield, 1998). Para la serie Sunspots, las entradas elegidas fueron los rezagos 1 al 4 (parámetros empleados por Cottrell *et al.* 1995).
- Se consideraron modelos que varían desde una hasta diez neuronas en la capa oculta.
- Cada modelo fue optimizado separadamente, con los algoritmos de regla delta generalizada y Rprop.
- Para cada modelo considerado (que se obtiene variando la cantidad de neuronas en la capa oculta), se obtuvo modelo con el menor error cuadrático medio, elegido tras repetir cada proceso 50 veces, donde cada punto de inicio se eligió de manera aleatoria.

RESULTADOS Y DISCUSIÓN

Resultados obtenidos Experimento 1: En la tabla 1, se exhiben la configuración de red neuronal usada en el experimento y los errores obtenidos con cada algoritmo de entrenamiento. Estos, igualmente, son exhibidos en la figura 2. En la figura 2a, se presenta el gráfico de la serie Airline y su respectivo ajuste con los algoritmos Regla Delta y RProp. La figura 2b, ilustra el mismo proceso para la serie Sunspots.

Las figuras 2a y 2b demuestran que los ajustes obtenidos por ambos algoritmos (regla delta y RProp), bajo iguales condiciones de la red neuronal, son altamente diferentes. Esto demuestra la incidencia que tiene en la práctica el método de entrenamiento utilizado.

Resultados obtenidos Experimento 2: En la figura 3, se grafican los respectivos MSE obtenidos con la ejecución del experimento para cada algoritmo de entrenamiento considerado versus la cantidad de neuronas en la capa oculta para la serie Airline (3a) y Sunspots (3b). En ambos paneles, se observa que el algoritmo de Regla Delta no es la mejor opción de entrenamiento, toda vez que el MSE obtenido siempre es mayor que el logrado usando RProp. Si bien, al aumentar el número de neuronas en la capa ocultase da un mejor ajuste a las muestras de calibración; el comportamiento teórico secuencialmente decreciente no es evidente siempre en éste método. Este comportamiento es opuesto a lo evidenciado teóricamente.

Por su parte, el algoritmo RProp permite encontrar errores secuencialmente decrecientes en apariencia; sin embargo, a partir de determinada neurona no existe una disminución notable en el error de ajuste, por lo que no se alcanza la meta de cero; incluso, manifestando aumentos leves en el error calculado.

La evaluación empírica del desempeño de los métodos de entrenamiento regla delta generalizada y Rprop permiten concluir que no se puede garantizar el cumplimiento del criterio de reducción del error de ajuste a medida que se aumenta la complejidad del modelo, adicionando neuronas en la capa oculta. Ambos algoritmos fallan al no presentar reducciones constantes en el error de ajuste y exhibir un equilibrio en

Tabla 1. Relación de Parámetros de Red y Errores de ajuste, para series Airline y Sunspots.

Serie	Entradas (P)	Neuronas Ocultas (H)	MSE Regla Delta	MSE RProp
Airline	1, 12, 13	2	0,0504	0,0035
SunSpots	1, 2, 3, 4	4	645,88	149,42

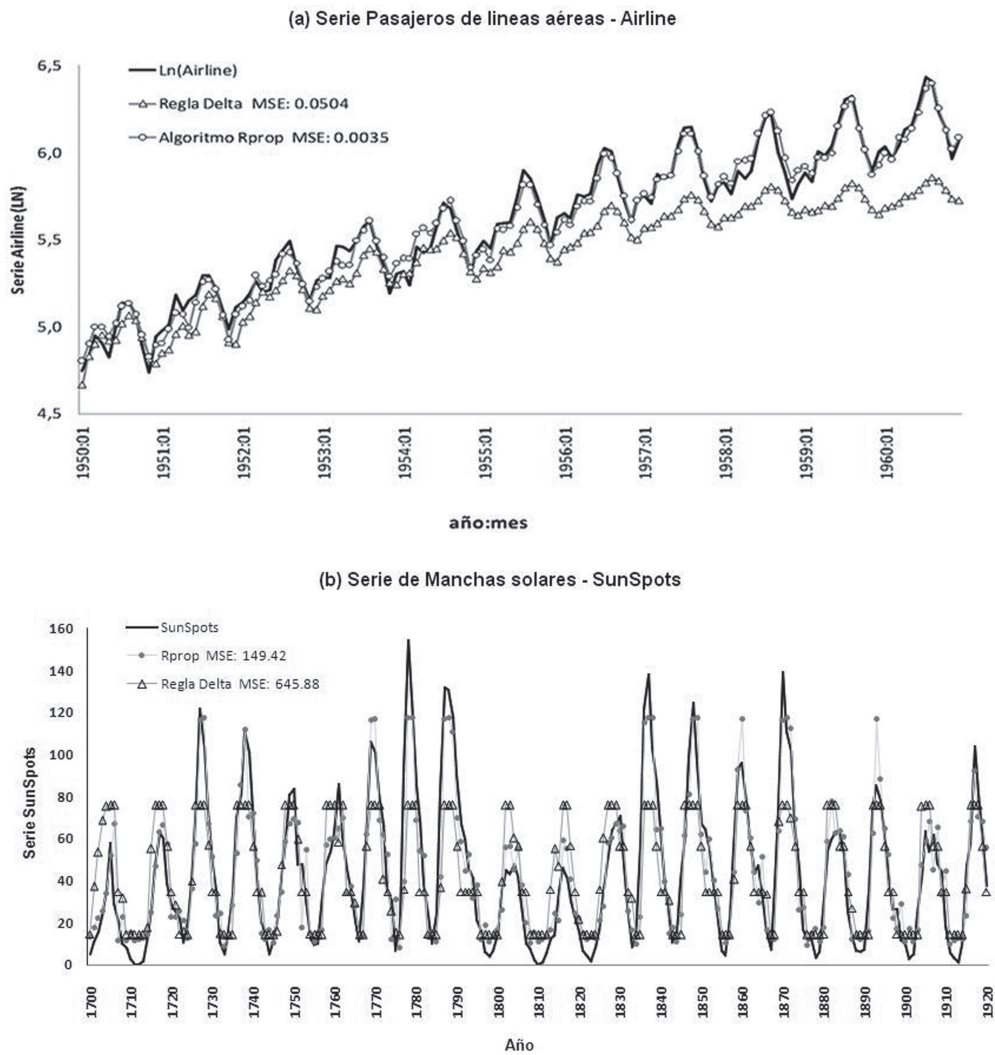


Figura 2. Ajuste de la serie Airline (a) y SunSpots (b), usando redes neuronales entrenadas con Regla Delta y Rprop.

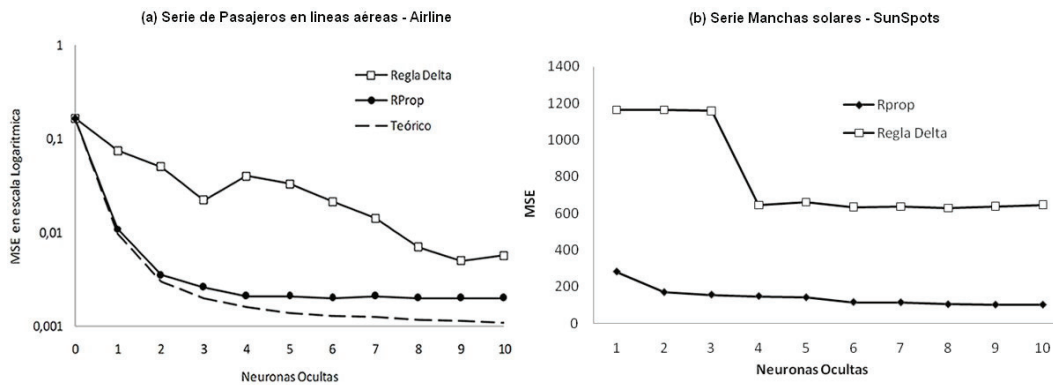


Figura 3. Comportamiento del MSE en serie Airline (a) y SunSpots (b), para modelos con entradas fijas y neuronas ocultas de 1 a 10.

un valor diferente a cero para el MSE, manifestando así, un comportamiento diferente al teóricamente planteado. Este resultado infiere, directamente, el proceso de construcción del modelo y, por ende, la selección adecuada del mejor modelo, toda vez que no se podría garantizar que no haya un modelo con mejor ajuste que el elegido.

Si bien, la falta de cumplimiento del supuesto se puede deber a otros factores, como implementación incorrecta del modelo de red neuronal, configuración de red inadecuada para las características de los datos o métodos de entrenamiento inadecuados, se descarta que dichos factores sean los causantes en este caso, puesto que, en primer lugar, los modelos de red neuronal adoptados son sugeridos como mejores modelos en los trabajos de Faraway & Chatfield (1998) y Cottrell *et al.* (1995); y en segundo lugar, se siguió un protocolo estándar para la especificación, la construcción y la selección del modelo, el cual, fue replicado una cantidad suficiente de veces, para descartar aleatoriedad en los resultados.

Implicaciones en la selección de entradas relevantes: La selección de las variables de entrada depende, en gran medida, del conocimiento que posea el modelador acerca de la serie de tiempo y es tarea de éste elegir, según algún criterio previamente fijado, la necesidad de cada variable dentro del modelo. La importancia primaria de hacer una selección adecuada radica en las dificultades de convergencia en el aprendizaje que puede acarrear el incluir entradas irrelevantes y el pobre desempeño del modelo.

Las metodologías para la selección de entradas relevantes, se basan en el supuesto que el algoritmo de optimización esta adecuadamente desarrollado, lo que degenera en consecuencias relacionadas con el incumplimiento de este supuesto, donde no se puede garantizar que la selección de las variables sea correcta y más aún que no se incluyan variables irrelevantes.

Implicaciones en la selección de neuronas ocultas: El incumplimiento del supuesto de reducción del error imposibilita un proceso constructivo de selección de variables, toda vez que, no se puede garantizar que el error disminuya al adicionar una nueva entrada. Más aún, esto degenera en las siguientes dificultades:

- Falta de convergencia de la red: el error no siempre alcanza el nivel deseado o el tiempo para hacerlo es alto.
- Pobre generalización: a menudo, algunos algoritmos, como el de la regla delta generalizada, presentan dificultades para replicar comportamientos, manifestando un amplio rango de valores del error, incluso, al comprender muchas variables.
- Selección errónea de variables: las dificultades en convergencia y generalización pueden conducir a la

inclusión de variables irrelevantes en el modelo, con el afán de alcanzar un nivel del error deseado y esto, a su vez, conduce a modelos altamente complejos (muchas variables).

Implicaciones en la construcción del modelo: El incumplimiento del supuesto de reducción del error influencia el proceso general de construcción del modelo, afectando el desempeño de la red en convergencia, capacidad de generalización o ajuste, selección adecuada de criterios de parada y robustez, toda vez que:

- No es posible, a partir de los resultados obtenidos, extraer reglas de comportamiento que nos conduzcan a una selección acertada de las variables y parámetros del modelo.
- No es posible seguir una metodología y proceso de modelado replicable.
- Conduce a una pobre generalización afectando, directamente, procesos orientados a la predicción, en la medida que se dificulta replicar comportamientos, siendo acentuado en patrones desconocidos.

Estos aspectos y otros tantos mencionados anteriormente, justifican estudios posteriores de la temática, orientados a la evaluación sistemática de diferentes métodos de entrenamiento, que permitan extraer tales reglas de comportamiento, con miras a una adecuada especificación del modelo de red neuronal.

Implicaciones metodológicas y conceptuales: El incumplimiento del supuesto de reducción del error, se puede ver como una explicación coherente de los resultados inconsistentes que, a menudo, se encuentran en la literatura acerca del desempeño de los modelos de redes neuronales. Las pruebas sobre los algoritmos de entrenamiento, se limitan a verificar la minimización del error, su capacidad de convergencia y generalización; sin embargo, aspectos como la verificación de los supuestos de reducción del error no son tenidos en cuenta a la hora de evaluar las bondades de su uso.

Si se garantiza la convergencia del error es posible encontrar redes neuronales con un mejor de ajuste y una mejor generalización, mientras que, usando los métodos tradicionales, se puede producir un estancamiento en cierto valor del error de tal forma que aunque se aumente la complejidad del modelo no es posible llegar a errores más bajos.

Conflictos de intereses: Este artículo fue preparado y revisado con la participación de todos los autores, quienes declaran que no existe ningún conflicto de intereses que ponga en riesgo la validez de los resultados presentados.

BIBLIOGRAFÍA

1. ADYA, M.; COLLOPY, F. 1998. How effective are neural networks at forecasting; prediction? A review; evaluation. *J. Forecasting (USA)*. 17:481-495.
2. ANASTASIADIS, A.D.; MAGOULAS, G.D.; VRAHATIS, M.N. 2003. An Efficient Improvement of the Rprop Algorithm. *Proceedings of the First International Workshop on Artificial Neural Networks in Pattern Recognition. University of Florence (ITALY)*. p.197-201.
3. COTTRELL, M.; GIRARD, B.; GIRARD, Y.; MANGEAS, M.; MÜLLER, C. 1995. Neural modeling for time series: a statistical stepwise method for weight elimination. *IEEE Transactions on Neural Networks (USA)*. 6(6):1355-1364.
4. CRONE, S.; KOURENTZES, N. 2009. Input-variable Specification for Neural Networks - An Analysis of Forecasting low and high Time Series Frequency. *Proceedings of the International Joint Conference on Neural Networks, IJCNN'09 (USA)*. p.619-626.
5. FAHLMAN, S. 1989. Faster-learning variations of back-propagation: An empirical study. En: Touretzky, D., Hinton, G.; Sejnowski, T. (eds) *Proceedings of the 1988 Connectionist Models Summer School (USA)*. p.38-51
6. FARAWAY, J.; CHATFIELD, C. 1998. Time series forecasting with neural networks: a comparative study using the airline data. *Appl. Statist. (USA)*. 47:231-250.
7. GHIASSI, M.; SAIDANE, H.; ZIMBRA, D.K. 2005. A dynamic neural network model for forecasting time series events. *International J. Forecasting (USA)*. 21:341-362.
8. HAGAN, M.T.; DEMUTH, H.B.; BEALE, M.H. 1996. *Neural Network Design*. Ed. PWS Publishing, Boston: MA(USA).
9. HAMILTON, J.D. 1994. *Time Series Analysis*. Princeton, New Jersey: Princeton University Press (USA). 820p.
10. HORNIK, K.; STINCHICOMBE, M.; WHITE, H. 1989. Multilayer Feedforward Networks are Universal Approximators. *Neural Networks (USA)*. 2(5):359-366.
11. MURATA, N.; YOSHIZAWA, S.; AMARI, S. 1994. Network information criterion-determining the number of hidden units for an artificial neural network model. *IEEE Transactions on Neural Networks (USA)*. 5:865-872.
12. QI, M.; ZHANG, P.G. 2001. An investigation of model selection criteria for neural network time series forecasting. *European J. Operational Research (NORWAY)*. 132:666-680.
13. TANG, Z.; KOEHLER, J.G. 1994. Deterministic global optimal FNN training algorithms. *Neural Networks (USA)*. 7:1405-1412.
14. VELÁSQUEZ, J.D.; DYNER, I.; SOUZA, R.C. 2008. Modelado del precio de la electricidad en brasil usando una red neuronal autorregresiva. *Ingeniare. Rev. Chilena Ingeniería*. 16(3):394-403.
15. ZHANG, P.G.; PATUWO, B.E.; HU, M.Y. 1998. Forecasting with artificial neural networks: the state of the art. *International J. Forecasting (USA)*. 14(1):35-62.
16. ZHANG, G.P.; PATUWO, B.E.; HU, M.Y. 2001. A simulation study of artificial neural networks for non linear time-series forecasting. *Computers & Operations Research (USA)*. 28(4):381-396.

Recibido: Noviembre de 12 de 2010

Aceptado: Marzo 28 de 2011