

Predicción de series temporales con redes neuronales: una aplicación a la inflación colombiana

Forecasting Time Series with Neural Networks: An Application to the Colombian Inflation

JUAN CAMILO SANTANA*

UNIVERSIDAD FEDERAL DE PERNAMBUCO, BRASIL

Resumen

Evaluar la capacidad de las redes neuronales en la predicción de series temporales es de sumo interés. Una aplicación que pronostique valores futuros de la serie de inflación colombiana permite mostrar que las redes neuronales pueden ser más precisas que las metodologías SARIMA de Box-Jenkins y el suavizamiento exponencial. Además, los resultados revelan que la combinación de pronósticos que hacen uso de las redes neuronales tiende a mejorar la capacidad de predicción.

Palabras Claves: Perceptron multicapas, modelos SARIMA, suavizamiento exponencial, combinación de pronósticos, componentes no observables.

Abstract

Evaluating the usefulness of neural network methods in predicting the Colombian Inflation is the main goal of this paper. The results show that neural networks forecasts can be considerably more accurate than forecasts obtained using exponential smoothing and SARIMA methods. Experimental results also show that combinations of individual neural networks forecasts improves the forecasting accuracy.

Key words: Multilayer perceptron, SARIMA models, Exponential smoothing, Combination of forecasts, Unobservable components.

*Maestro en Estadística. E-mail: csantana@cable.net.co

1. Introducción

Una estrategia alternativa que utilice redes neuronales será considerada con el objetivo de elaborar pronósticos sobre la serie de inflación colombiana, es decir, sobre las variaciones del índice de precios al consumidor (IPC), publicado mensualmente por el Dane. El desarrollo de metodologías que permitan pronosticar y comprender el comportamiento de la inflación es de sumo interés para muchos sectores de la población y la economía. De esta forma, su cuantificación resulta necesaria para la toma de decisiones dentro del contexto económico y social.

Investigaciones y aplicaciones de las redes neuronales a nivel económico, en el ámbito colombiano, han sido realizadas por Misas, López & Borrero (2002), Misas, López, Arango & Hernández (2003) y recientemente Jalil & Misas (2006) y Aristizábal & Misas (2006), quienes han evidenciado las bondades de las redes neuronales en la predicción, comparadas con otras metodologías tradicionales. Motivados por las características de las redes neuronales, nuestra idea principal es describir una metodología alternativa, referente al mecanismo de modelamiento de las redes neuronales y diferente a la expuesta en artículos preliminares en el contexto colombiano, que permita unirse al abanico de técnicas ya existentes en esta línea. El conocimiento de la inflación (información que es publicada mensualmente) resulta fundamental al tomar decisiones tanto de control sobre los instrumentos de política monetaria, como tasas de interés, encajes bancarios u operaciones de diversa índole que aumenten o contraigan la disponibilidad de recursos en la economía. Por ejemplo, para los mercados de capitales es fundamental contar con políticas de los bancos centrales enfocadas a mantener bajo control la inflación, en la medida en que se aminora la incertidumbre de los agentes, las tasas de interés pueden reducirse, y por ende, se estimula la actividad económica.

Las predicciones obtenidas de la inflación con redes neuronales serán contrastadas con las obtenidas a través de las metodologías SARIMA de Box-Jenkins y el suavizamiento exponencial, como también de la combinación de pronósticos. La utilidad de estas metodologías tradicionales en el pronóstico es analizada por Ospina & Zamprogno (2003), quienes evalúan el desempeño de ciertas técnicas en la predicción de series temporales.

Hornik et al. (1989) y Cybenko (1989), entre otros, han demostrado que las redes neuronales son aproximadores universales y que el perceptron multicapas es una de las arquitecturas más utilizadas en la solución de problemas debido a su fácil uso y aplicabilidad; véase Cohen et al. (1993), Narendra & Parthasaranty (1990) y Wiegand et al. (1990) para aplicaciones con redes neuronales. Dentro del área estadística, las redes neuronales son consideradas como métodos no lineales, no paramétricos y multivariados (véase Zhang et al. 1998).

Para evaluar el desempeño de las redes neuronales en el pronóstico, se utilizarán las metodologías de Box-Jenkins y suavizamiento exponencial, al igual que la combinación de pronósticos. Se supondrá que el lector tiene un conocimiento básico con relación a cada tema. No obstante, el desarrollo metodológico de la teoría de Box-Jenkins puede ser estudiado más ampliamente en Box & Jenkins (1976), Box et al. (1994) y Morettin & Toloí (2004), entre otros; para más detalles

sobre los algoritmos de suavizamiento exponencial, véase Morettin & Toloí (2004) y Montgomery & Johnson (1976), y con relación a la combinación de pronósticos, Barnard (1963) y Hendry & Clements (2004) hacen desarrollos completos en este sentido. Igualmente, el conocimiento sobre la obtención de componentes no observables, como tendencia y estacionalidad, a través de la metodología de Box-Jenkins será importante en el desarrollo de este artículo. El lector podrá referirse a Maravall & Kaiser (2000) para una discusión mayor sobre este tema.

El presente artículo se encuentra organizado de la siguiente forma: en la sección 2 se discuten las principales características de la red neuronal perceptron multicapas. En la sección 3 se calculan pronósticos para la inflación; se realizan comparaciones con las metodologías clásicas y la combinación de pronósticos. Por último, en la sección 4, se presentan las principales conclusiones.

2. Redes neuronales artificiales

La arquitectura de redes neuronales más ampliamente utilizada es la que se conoce con el nombre de perceptron multicapas, la cual se caracteriza por el hecho de que sus neuronas se agrupan en capas por niveles. Cada una de estas capas está constituida por un conjunto de neuronas. Hay tres tipos de capas diferentes: la capa de entrada, las capas ocultas y la capa de salida, como se observa en la figura 1.

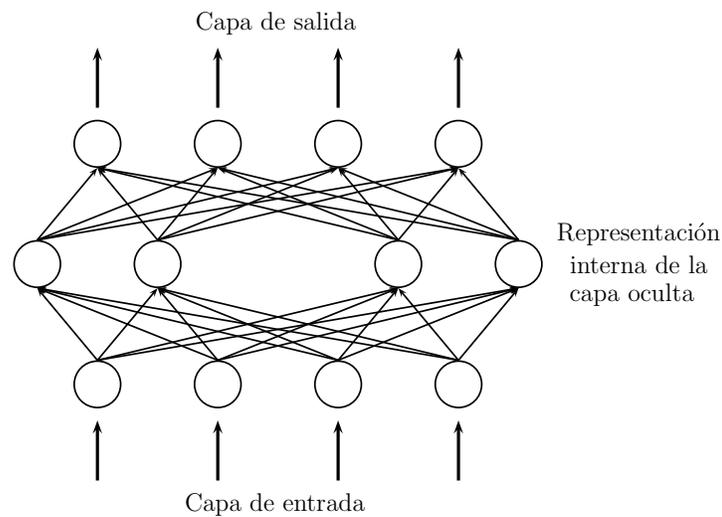


FIGURA 1: Red neuronal *feedforward*.

Las neuronas de la capa de entrada se encargan únicamente de recibir señales o patrones que vienen del exterior y propagan tales señales a todas las neuronas de la capa siguiente. La última capa actúa como salida de la red, proporcionando

al exterior la respuesta de la red para cada uno de los patrones de entrada. Las neuronas de las capas ocultas realizan un procesamiento no lineal de los patrones recibidos. Como se observa en la figura 1, las conexiones del perceptron multicapas están siempre dirigidas hacia delante, i. e., las neuronas de una capa se conectan con las neuronas de la capa siguiente; por tal motivo reciben el nombre de redes alimentadas hacia delante o redes *feedforward*. A las conexiones se les asocia un número real llamado peso de la conexión y a las neuronas de la red un umbral, que en el caso del perceptron multicapas es tratado como una conexión adicional a la neurona.

2.1. Propagación de los patrones de entrada

Una vez descrita la forma como fluye la información a través de la arquitectura del perceptron multicapas, presentaremos a continuación las expresiones para el cálculo de las activaciones de las neuronas de la red.

Considere un perceptron multicapas con C capas ($C - 2$ capas ocultas) y n_c neuronas en la capa c , para $c = 1, 2, \dots, C$. Sea $W^c = (w_{ij}^c)$ la matriz de pesos asociada a las conexiones de la capa c a la capa $c + 1$, para $c = 1, 2, \dots, C - 1$, en que w_{ij}^c representa el peso de la conexión de la neurona i de la capa c a la neurona j de la capa $c + 1$; además, sea $U^c = (u_i^c)$ el vector de umbrales de las neuronas de la capa c para $c = 2, \dots, C$. Es denotada por a_i^c la activación de la neurona i de la capa c ; estas activaciones se calculan de la siguiente forma:

$$a_i^c = f \left(\sum_{j=1}^{n_{c-1}} w_{ij}^{c-1} a_j^{c-1} + u_i^c \right)$$

para $i = 1, 2, \dots, n_c$ y $c = 2, 3, \dots, C$. Las activaciones para la primera capa corresponden simplemente con las observaciones de entrada a la red.

La función $f(\cdot)$ es llamada *función de activación o transferencia*. Para el perceptron multicapas, las funciones de activación más utilizadas son la logística o sigmoide y la tangente hiperbólica; sin embargo, también se utilizan otras funciones de activación (véase Gately 1996). El propósito de la función de activación o transferencia es no permitir la salida de valores muy grandes, los cuales pueden retrasar el proceso de convergencia del algoritmo de entrenamiento o aprendizaje, que se describirá a continuación.

2.2. Algoritmo de retropropagación

El algoritmo de aprendizaje es el mecanismo mediante el cual se van adaptando y modificando todos los parámetros de la red. El problema de aprendizaje de la red es un problema de minimización de la siguiente forma:

$$\min_W E$$

W es el conjunto de parámetros de la red (pesos y umbrales) y E una función del error que evalúa la diferencia entre la salida de la red y la salida deseada. En la

mayoría de los casos la función del error es definida por:

$$E = \frac{1}{N} \sum_{n=1}^N e(n) \tag{1}$$

donde N es el número de observaciones o patrones y $e(n)$ es el error cometido por la red para el n -ésimo patrón, que es dado por:

$$e(n) = \frac{1}{2} \sum_{i=1}^{n_c} (s_i(n) - y_i(n))^2 \tag{2}$$

con $Y_n = (y_1(n), y_2(n), \dots, y_{n_c}(n))$ y $S_n = (s_1(n), s_2(n), \dots, s_{n_c}(n))$, los vectores de salida de la red y salida deseada para el n -ésimo patrón, respectivamente.

De esta forma, si W^* es un mínimo de la función (1), i. e., el punto donde el error es el menor posible y la salida de la red es próxima de la deseada, se obtiene el fin del proceso de aprendizaje.

Para el perceptron multicapas, el método de optimización no lineal más utilizado es el *steepest descent* sobre la función E . De esta forma, cada parámetro w de la red es modificado para cada patrón de entrada n de acuerdo con la siguiente ley de aprendizaje:

$$w(n) = w(n - 1) - \eta \frac{\partial e(n)}{\partial w} \tag{3}$$

con $e(n)$ definido en (2) y η la tasa de aprendizaje que influye en la magnitud de desplazamiento sobre la superficie del error. El método del gradiente puede ser aplicado de forma eficiente, resultando en el conocido algoritmo de retropropagación o regla *delta generalizada* (véase Rumelhart et al. (1986b) e Isasi & Galván (2004) para una descripción más detallada sobre esta regla).

Como puede observarse en (3), el cambio en un peso es proporcional al gradiente del error, con la proporcionalidad dada por el parámetro η . Valores altos de la tasa de aprendizaje en principio podrían favorecer una convergencia con menos iteraciones, pues permite avanzar rápidamente en la superficie del error. Sin embargo, tasas de aprendizaje altas pueden tener consecuencias negativas sobre el aprendizaje, haciendo que el método salte u oscile alrededor del mínimo. Valores pequeños de las tasas de aprendizaje pueden evitar estos problemas, aunque posiblemente lleven a una convergencia más lenta del algoritmo de aprendizaje, debido a que la magnitud del desplazamiento sobre la superficie del error es menor.

Un método simple para evitar la inestabilidad en el algoritmo de aprendizaje, debido a la tasa de aprendizaje, consiste en modificar (3) a través de la inclusión de un segundo término llamado *momento*, obteniendo de esta forma la siguiente ley:

$$w(n) = w(n - 1) - \eta \frac{\partial e(n)}{\partial w} + \alpha \Delta w(n - 1) \tag{4}$$

donde α es un número positivo que actúa como ponderador. Esta regla fue propuesta por Rumelhart et al. (1986a) y preserva las propiedades de la regla definida en (3), en el sentido en que modifica los parámetros de la red para minimizar la

función del error (1). El nuevo término, $\alpha\Delta w(n-1)$, incorpora al método alguna inercia, haciendo que la modificación actual del parámetro dependa sólo de la dirección de la modificación anterior y consigue evitar oscilaciones. Haciendo cálculos sucesivos sobre $\Delta w(n-1)$, Isasi & Galván (2004) exhiben una expresión más general de (4):

$$w(n) = w(n-1) - \eta \sum_{t=0}^n \alpha^{n-t} \frac{\partial e(t)}{\partial w} \quad (5)$$

El proceso de aprendizaje del perceptron multicapas debe ser finalizado cuando $\frac{\partial E}{\partial w} \approx 0$, momento en el cual los parámetros de la red no cambian de forma perceptible entre iteraciones consecutivas.

2.3. Capacidad de generalización

A la hora de evaluar el comportamiento de la red, y en particular del perceptron multicapas, no importa saber si la red aprendió con éxito los patrones utilizados durante el aprendizaje, sino conocer el comportamiento de la red frente a patrones que no fueron utilizados durante el entrenamiento.

Para tal fin, es necesario disponer de dos conjuntos de patrones: el *conjunto de entrenamiento*, que entrena y modifica los pesos y umbrales de la red, y el *conjunto de validación*, que mide la capacidad de la red para responder correctamente a los patrones que no fueron ingresados durante el entrenamiento. Cuando la red aproxima correctamente los patrones de aprendizaje, pero no responde bien a los patrones de validación, se dice que hubo subaprendizaje de la red, posiblemente ocasionado por varios factores, como el uso de un número excesivo de neuronas o capas ocultas, implicando un aumento en el número de parámetros a ser estimados; véase Isasi & Galván (2004), Kaastra & Boyd (1996) y Zhang et al. (1998) para una discusión mayor sobre el tema.

3. Aplicación

El análisis referente al modelamiento y predicción que se presenta en seguida se basa en el índice mensual de precios al consumidor (IPC). La serie de variaciones mensuales (inflación) es calculada por el Dane y se puede consultar a través de la página www.dane.gov.co. El período comprendido entre enero de 1998 y diciembre de 2005 se utiliza para el modelamiento y pronóstico. Adicionalmente, la transformación logaritmo se usa para controlar varianza.

Utilizamos el período comprendido entre enero de 1998 a junio de 2005 para el modelamiento de la serie con la finalidad de obtener las predicciones del período de julio a diciembre de 2005. Nuestro principal interés es observar cómo cada modelo captura la estructura dinámica de la serie y la refleja sobre el período de predicción. La razón por la cual se consideran seis observaciones en la predicción y no un período más largo, como un año, radica en la creencia de que la existencia

de cambios estructurales puede hacer que el modelo no recoja adecuadamente la dinámica que exhibiría la serie verdadera y, por ende, se podrían presentar elevados errores de pronóstico. De esta forma, conocer la dinámica de la inflación del primer trimestre del año puede mejorar la capacidad predictiva del modelo y permitir una mejor predicción para el segundo semestre.

Respecto a las predicciones realizadas por cada metodología se adoptan dos formas de medición del error de predicción para establecer las comparaciones:

- **Error cuadrático medio (MSE).** Si Z_1, Z_2, \dots, Z_t son las observaciones de la serie temporal y $Z_{t+1}, Z_{t+2}, \dots, Z_{t+h}$ sus h predicciones respectivas, entonces:

$$MSE = \frac{1}{h} \sum_{k=1}^h e_{t+k}^2, \text{ donde } e_{t+k} \text{ es el error de predicción de } Z_{t+k}, \text{ el valor de la serie en el instante } t+k.$$

- **Error absoluto medio (MAE).** Aquí,

$$MAE = \frac{1}{h} \sum_{k=1}^h |e_{t+k}|, \text{ donde } e_{t+k} \text{ es el error de predicción en el instante } t+k, k = 1, 2, \dots, h.$$

La serie de inflación para el período comprendido entre enero de 1998 y diciembre de 2005 se presenta en la figura 2, en la cual es posible observar fuertes indicios de no estacionariedad y estacionalidad.

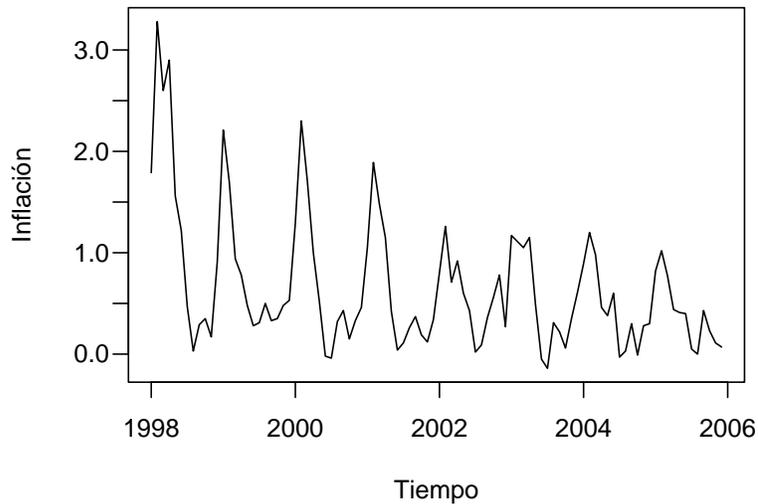


FIGURA 2: Serie de inflación: enero/98-diciembre/05.

3.1. Modelamiento y pronóstico

Los algoritmos de suavizamiento exponencial aditivo y multiplicativo de Holt-Winters se utilizaron para obtener, inicialmente, predicciones referentes al período que se extiende de enero de 1998 a junio de 2005. Los valores optimizados de las constantes de suavizamiento resultaron semejantes para los modelos aditivo y multiplicativo, en este caso, consecuencia de utilizar la transformación logaritmo sobre la serie; igualmente, las predicciones obtenidas a través de estos dos modelos resultaron semejantes, así que el modelo aditivo fue escogido para la generación de las predicciones. La tabla 1 exhibe las respectivas predicciones un paso ($h = 1$) y seis pasos ($h = 6$) al frente, sobre la escala original.

TABLA 1: Predicción de la inflación a través del suavizamiento exponencial aditivo.

Período	$h = 1$	$h = 6$
Jul-05	-0.113	-0.113
Ago-05	0.067	0.057
Sep-05	0.230	0.224
Oct-05	0.037	0.020
Nov-05	0.316	0.286
Dic-05	0.340	0.323

Para el proceso de identificación del modelo a través de la metodología SARIMA de Box-Jenkins se utilizó el criterio de información Bayesiano (BIC). El programa Tramo-Seats se usó para estimar los parámetros del modelo SARIMA. Este programa es gratuito y puede ser obtenido a través de la página web del Banco de España (www.bde.es).

El modelo $SARIMA(1, 1, 1) \times (0, 1, 1)_{12}$ fue escogido para la serie de la inflación. El análisis de diagnóstico permite observar que no hay evidencias contra la hipótesis de ausencia de autocorrelación de los residuos, como tampoco contra la hipótesis de normalidad de los residuos al nivel de significancia del 1%.

La tabla 2 exhibe las estadísticas del análisis de diagnóstico sobre los residuales del modelo seleccionado, donde $SE(Res)$ es el error estándar de los residuos; $Q-Val$ es la estadística de Ljung-Box para probar la hipótesis de correlación serial, calculada sobre 24 autocorrelaciones (en todos los casos se utiliza la distribución asintótica χ^2 , con 21 grados de libertad); $N-test$ es la estadística de Bowman-Shenton para probar la hipótesis de normalidad (la distribución asintótica $\chi^2(2 \text{ g.l.})$ es utilizada); $SK(t)$ es la estadística que se usa para probar si la asimetría es cero contra si es diferente de cero; $KUR(t)$ es la estadística utilizada para probar si el exceso de kurtosis es cero contra si es diferente de cero; $Q2$ es la estadística de McLeod & Li (1983) para probar la linealidad del proceso (la distribución asintótica $\chi^2(24 \text{ g.l.})$ es utilizada); por último, $RUNS$ es la estadística que se usa para probar la hipótesis nula sobre aleatoriedad del conjunto de residuos. Todas las pruebas de hipótesis se realizan al nivel de significancia de 1%.

La ecuación del modelo estimado para la serie transformada de la inflación,

TABLA 2: Estadísticas de diagnóstico.

Serie	$SARIMA(1, 1, 1) \times (0, 1, 1)_{12}$
$SE(Res)$	0.1736
BIC	-3.3306
$Q - Val$	34.1100
$N - test$	1.3200
$SK(t)$	1.0600
$KUR(t)$	-0.4400
$Q2$	10.9600
$RUNS$	-0.2300
<hr/>	
	$\chi^2_2(99\%) = 9.21$
	$\chi^2_{21}(99\%) = 38.93$
	$\chi^2_{24}(99\%) = 42.98$

denotada por Y_t , es dada por:

$$[t : -3.4] \quad [t : -66.0] \quad [t : -3.9] \\ (1 + 0.429B)Y_t = (1 + 0.991B)(1 + 0.769B^{12})a_t$$

donde:

$$Y_t = (1 - B)(1 - B^{12})Z_t$$

y

$$a_t \sim N(0, (0.1736)^2)$$

Las componentes de tendencia y estacionalidad servirán para el modelamiento con redes neuronales y fueron extraídas del modelo SARIMA (véase Gómez & Maravall (1996) para una discusión mayor sobre el tema). La figura 3 exhibe las dos componentes no observables entre enero de 1998 y junio de 2005. Note que la tendencia exhibida por la inflación sigue una trayectoria decreciente desde finales de 2000 hasta mitad de 2005. Con la componente estacional podemos observar que en los meses de febrero se registran los datos de inflación más altos del año, y en julio los más bajos.

La tabla 3 contiene las predicciones un paso ($h = 1$) y seis pasos ($h = 6$) al frente, utilizando el modelo SARIMA estimado anteriormente.

Para el modelamiento de la inflación a través de redes neuronales se utilizó la propuesta de Varfis & Versino (1990). Cada serie es reescalada en el intervalo $[-1, 1]$ antes de incluir variables rezagadas, componentes de tendencia y estacionalidad como neuronas de entrada. Aunque otras transformaciones de reescalamiento han sido propuestas con el objetivo de acelerar el proceso de entrenamiento (véase Zhang et al. 1998), intentamos ser consistentes con la función de activación seleccionada. La función de activación utilizada es la tangente hiperbólica antisimétrica descrita en Haykin (1994). Adicionalmente, se consideran 2 capas ocultas, 2 parámetros de aprendizaje, 2 parámetros de momento, 10000 *epochs* y un máximo

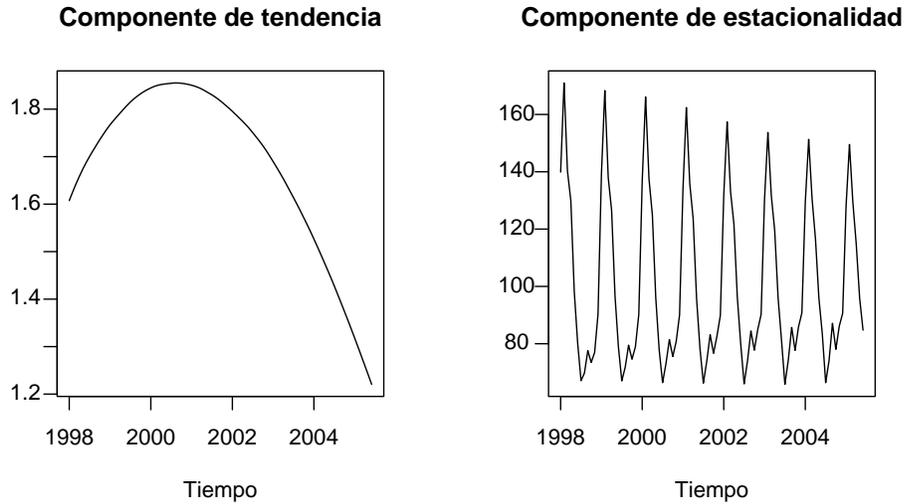


FIGURA 3: Componentes no observables de la inflación.

TABLA 3: Predicción de la inflación a través de la metodología de Box-Jenkins.

Período	$h = 1$	$h = 6$
Jul-05	0.044	0.044
Ago-05	0.170	0.164
Sep-05	0.200	0.281
Oct-05	0.208	0.150
Nov-05	0.356	0.318
Dic-05	0.309	0.410

de 6 neuronas por capa oculta, siguiendo las sugerencias de Kaastra & Boyd (1996) y Zhang et al. (1998). Tres clases de arquitectura de redes neuronales se definen:

- a) Red 1. Una capa oculta es considerada, con un número máximo de 6 neuronas.
- b) Red 2. Dos capas ocultas son consideradas, con igual número de neuronas en cada capa oculta, con máximo 6 neuronas.
- c) Red 3. Dos capas ocultas son consideradas, con una neurona adicional en la segunda capa y cada una con un número máximo de 6 neuronas.

El objetivo principal fue obtener redes con buen ajuste y la mejor predicción posible. Para bautizar las redes neuronales identificadas, se utilizó la notación de Souza & Zandonade (1993), dada por $ANN(n_1, n_2, \dots, n_C)$, donde n_c , $c = 1, 2, \dots, C$ es el número de neuronas en la capa c . De esta forma, establecido un conjunto fijo de neuronas de entrada, 72 redes neuronales fueron simuladas, es decir, 24 redes

neuronales por cada tipo de red. Consideramos adicionalmente la simulación de un conjunto de redes neuronales eximiendo la componente de tendencia, de tal forma que pudiéramos evaluar empíricamente la ventaja de introducir tal componente.

Un programa escrito en lenguaje R (R Development Core Team 2005) se usó para el proceso de simulación y predicción con redes neuronales. En promedio, cada red neuronal utilizó 10 minutos durante el proceso de entrenamiento, para un total de 12 horas de simulación. Las variables de entrada a la red se determinaron por medio de un modelo autorregresivo.

3.2. Resultado de las simulaciones

La red $ANN(17, 5, 1)$ se identificó con 17 neuronas de entrada definidas por el primero, segundo, octavo, noveno y duodécimo rezagos de la serie y 12 variables *dummy* que identifican cada mes del año. Esta red es del tipo 1 con $\eta = 0.1$, $\alpha = 0.1$ y es la que proporcionó la mejor predicción un paso al frente tanto en *MSE* como en *MAE*. La red $ANN(15, 4, 1)$, con variables rezagadas de primero, duodécimo, decimotercer orden y 12 variables *dummy*, exhibió las mejores predicciones múltiples pasos según *MSE*. Esta red es del tipo 1, con $\eta = 0.01$, $\alpha = 0.5$. En función del *MAE* las mejores predicciones múltiples pasos fueron proporcionadas por la red $ANN(17, 2, 2, 1)$ con rezagos de primero, segundo, octavo, noveno y duodécimo orden, junto con 12 variables *dummy* como antes. Ésta es una red tipo 2 con $\eta = 0.01$, $\alpha = 0.1$.

La tabla 4 exhibe las predicciones uno y múltiples pasos al frente para estas tres redes. Note que las tres redes encontradas anteriormente no incluyen la tendencia entre sus neuronas de entrada.

TABLA 4: Predicciones *ANN*: uno y seis pasos al frente.

Período	$ANN(17, 5, 1)$		$ANN(15, 4, 1)$		$ANN(17, 2, 2, 1)$	
	$h = 1$	$h = 6$	$h = 1$	$h = 6$	$h = 1$	$h = 6$
Jul-05	0.008	0.008	-0.013	-0.013	-0.063	-0.063
Ago-05	-0.116	-0.098	0.151	0.133	0.028	0.021
Sep-05	0.379	0.379	0.226	0.217	0.257	0.293
Oct-05	0.430	0.513	0.285	0.132	0.340	0.201
Nov-05	0.107	0.288	0.209	0.184	0.131	0.186
Dic-05	0.236	-0.030	0.238	0.234	0.344	0.345

Una vez obtenidas las predicciones a través de cada metodología, inspeccionaremos las medidas de error de predicción *MSE* y *MAE* para determinar cuál metodología, de forma individual, es la que proporciona los mejores pronósticos. La tabla 5 exhibe estos resultados; se destaca que las redes neuronales presentan las mejores predicciones tanto en un paso como en múltiples pasos, comparadas con metodologías tradicionales como el suavizamiento de Holt-Winters y SARIMA de Box-Jenkins.

La media ponderada de las predicciones individuales proporcionadas por el sua-

TABLA 5: Medidas de error de predicción.

Modelo	<i>MSE</i>		<i>MAE</i>	
	$h = 1$	$h = 6$	$h = 1$	$h = 6$
<i>HW</i>	0.037	0.035	0.183	0.178
<i>SARIMA</i>	0.033	0.036	0.152	0.158
<i>ANN</i> (17, 5, 1)	0.014	0.023	0.096	0.125
<i>ANN</i> (15, 4, 1)	0.018	0.018	0.123	0.124
<i>ANN</i> (17, 2, 2, 1)	0.022	0.019	0.120	0.108

vizamiento exponencial, por el método *SARIMA* de Box-Jenkins y por las redes neuronales, fue utilizada para obtener una predicción mejorada. Las ponderaciones fueron escogidas siendo inversamente proporcionales al error de predicción individual. Es importante resaltar que sólo discutiremos las combinaciones que proporcionaron las mejores predicciones según el *MSE*, es decir, aquellas combinaciones que proporcionaron el mínimo *MSE*. Para el estudio de estas combinaciones, un conjunto adicional de 16 redes neuronales con las mejores predicciones según el *MSE* fueron obtenidas a través de simulaciones, evaluando diferentes configuraciones de la capa de entrada.

Poseemos en total 21 modelos: el mejor modelo *SARIMA* según el *BIC*, el mejor modelo de Holt-Winters aditivo y 19 redes neuronales con el mejor ajuste y predicción posible. Se consideraron, entonces, 2^{21} posibles combinaciones y para cada combinación se calculó el *MSE*. El proceso de simulación para esta fase tuvo una duración de 8 minutos.

En la predicción un paso al frente, la combinación de las redes neuronales *ANN*(18, 2, 3, 1), *ANN*(17, 5, 1), *ANN*(17, 5, 5, 1) y *ANN*(17, 3, 1), que llamaremos *COMB*₁, exhibió las mejores predicciones con *MSE* = 0.005 y *MAE* = 0.058. Estos resultados fueron mejores que los registrados en la tabla 5. Para la predicción seis pasos al frente, las redes neuronales *ANN*(18, 2, 3, 1), *ANN*(14, 6, 6, 1), *ANN*₂(15, 4, 1) (diferente a *ANN*(15, 4, 1)) y *ANN*(17, 3, 1), que llamaremos *COMB*₂, presentaron los mejores resultados con *MSE* = 0.009 y *MAE* = 0.072.

La tabla 6 contiene las predicciones uno y seis pasos al frente, resultado de estas combinaciones, y la tabla 7 contiene las medidas de error de predicción, las cuales pueden ser contrastadas con las exhibidas en la tabla 5.

Las variables rezagadas y los coeficientes η y α de las redes neuronales que hicieron parte de estas combinaciones se describen en la tabla 8.

La combinación de pronósticos con redes neuronales, incluyendo las predicciones del modelo *SARIMA* y el suavizamiento de Holt-Winters, dejó a la combinación de las predicciones de los modelos *SARIMA*, *HW*, *ANN*(18, 2, 3, 1), *ANN*(17, 5, 1) y *ANN*(17, 5, 5, 1) (*COMB*₃) con la mejor predicción un paso al frente según el *MSE* y a la combinación de las predicciones de los modelos *SARIMA*, *HW*, *ANN*(18, 2, 3, 1), *ANN*(14, 6, 6, 1), *ANN*₂(15, 4, 1) y *ANN*(17, 3, 1) (*COMB*₄) con la mejor predicción múltiples pasos al frente según el *MSE*.

TABLA 6: Combinación de predicciones.

Período	$COMB_1$		$COMB_2$	
	$h = 1$	$h = 6$	$h = 1$	$h = 6$
Jul-05	0.008	0.008	0.019	0.019
Ago-05	0.050	0.032	0.045	0.030
Sep-05	0.313	0.329	0.221	0.275
Oct-05	0.203	-0.007	0.199	0.275
Nov-05	0.098	0.097	-2.251	0.096
Dic-05	0.171	0.623	0.194	0.224

TABLA 7: Medidas de error de predicción.

Combinación	MSE		MAE	
	$h = 1$	$h = 6$	$h = 1$	$h = 6$
$COMB_1$	0.005	0.063	0.058	0.163
$COMB_2$	0.940	0.009	0.467	0.072

TABLA 8: Redes neuronales utilizadas.

Red	Rezagos	η	α
$ANN(17, 5, 1)$	$z_1, z_2, z_8, z_9, z_{12}, 12D$	0.10	0.1
$ANN(15, 4, 1)$	$z_1, z_{12}, z_{13}, 12D$	0.01	0.5
$ANN(17, 2, 2, 1)$	$z_1, z_2, z_8, z_9, z_{12}, 12D$	0.01	0.1
$ANN(18, 2, 3, 1)$	$z_1, z_2, z_8, z_9, z_{12}, \text{Ten}, 12D$	0.10	0.1
$ANN(17, 5, 5, 1)$	$z_1, z_2, z_8, z_9, z_{12}, 12D$	0.01	0.1
$ANN(14, 6, 6, 1)$	$z_1, \text{Ten}, 12D$	0.01	0.5
$ANN_2(15, 4, 1)$	$z_1, z_{12}, z_{13}, 12D$	0.01	0.1
$ANN(17, 3, 1)$	$z_1, z_2, z_{12}, z_{13}, \text{Ten}, 12D$	0.01	0.5

Ten es la componente de tendencia y 12D las doce variables *dummy*.

Note que las redes neuronales consideradas en estas combinaciones se analizaron en combinación antes, proporcionando las mejores predicciones. Adicionalmente, la combinación de las predicciones de los modelos *SARIMA* y *HW* con las redes neuronales, redujo la *MSE* y el *MAE* para $COMB_3$ en la predicción múltiples pasos, comparada con los resultados obtenidos por $COMB_1$. Lo inverso ocurrió con $COMB_4$ en beneficio de las predicciones un paso al frente en comparación con $COMB_2$. La tabla 9 exhibe los resultados comentados anteriormente, y las predicciones obtenidas por $COMB_3$ y $COMB_4$ se presentan en la tabla 10.

TABLA 9: Medidas de error de predicción.

Combinación	<i>MSE</i>		<i>MAE</i>	
	$h = 1$	$h = 6$	$h = 1$	$h = 6$
$COMB_3$	0.007	0.038	0.066	0.134
$COMB_4$	0.127	0.011	0.217	0.084

TABLA 10: Combinación de predicciones.

Período	$COMB_3$		$COMB_4$	
	$h = 1$	$h = 6$	$h = 1$	$h = 6$
Jul-05	0.028	0.028	0.030	0.030
Ago-05	0.059	0.041	0.054	0.038
Sep-05	0.292	0.312	0.219	0.270
Oct-05	0.200	0.054	0.196	0.313
Nov-05	0.092	0.090	0.940	0.089
Dic-05	0.197	0.495	0.220	0.252

4. Conclusiones

Los resultados obtenidos ilustraron el uso de las redes neuronales en la predicción de series temporales. Un ejemplo aplicado sobre las variaciones en el índice de precios al consumidor (IPC) permitió observar que las predicciones obtenidas a través de redes neuronales tienden a ser más precisas que las originadas por metodologías tradicionales, como el suavizamiento exponencial y el método SARIMA de Box-Jenkins. Adicionalmente, la utilidad de las redes neuronales, en combinación con otras redes o metodologías tradicionales, se mostró efectiva en el proceso de predicción en términos del *MSE*. Finalmente, las redes neuronales sin la neurona de tendencia, en la mayoría de los casos exhibieron las menores medidas de error en la predicción, en comparación con la metodología utilizada que propone la tendencia como una neurona de entrada. En todos los casos se estimaron y seleccionaron redes que tuvieran el mejor ajuste posible sobre el conjunto de entrenamiento y la menor medida de error de predicción sobre el conjunto de validación con el fin de evitar el problema de subaprendizaje.

Recibido: febrero de 2006

Aceptado: abril de 2006

Referencias

- Aristizábal, M. & Misas, M. (2006), Evaluación asimétrica de una red neuronal artificial: una aplicación al caso de la inflación en Colombia, Technical report, Working Paper 377. Subgerencia de estudios económicos–Banco Central de Colombia.

- Barnard, G. A. (1963), 'New Methods of Quality Control', *Journal of the Royal Statistical Society A* **126**, 255–259.
- Box, G. E. P. & Jenkins, G. M. (1976), *Time Series Analysis: Forecasting and Control*, San Francisco: Holden-Day.
- Box, G. E. P., Jenkins, G. M. & Reinsel, G. (1994), *Time Series Analysis: Forecasting and Control*, 3a. edn, Englewood Cliffs: Prentice Hall.
- Cohen, M., Franco, H., Morgan, N., Rumelhart, D. & Abrash, V. (1993), *Advances in Neural Information Processing Systems*, Morgan Kaufmann, chapter Context-dependent multiple distribution phonetic modeling with MLPs, pp. 649–657.
- Cybenko, M. (1989), 'Aproximation by Superposition of a Sigmoidal Function', *Mathematics of Control, Signals and Systems* **2**, 303–314.
- Gately, E. (1996), *Neural Networks for Financial Forecasting*, New York: John Wiley and Sons.
- Gómez, V. & Maravall, A. (1996), Programs Tramo (Time series Regression with Arima noise, Missing observations, and Outliers) and Seats (Signal Extraction in Arima Time Series). Instructions for the User, Technical report, Working Paper 9628. Servicio de estudios-Banco de España.
- Haykin, S. (1994), *Neural Networks*, New York: McMillan College Publishing Company.
- Hendry, D. F. & Clements, M. P. (2004), 'Pooling of Forecast', *Econometrics Journal* **7**, 1–31.
- Hornik, K., Stinchcombe, M. & White, H. (1989), 'Multilayer Feedforward Networks and Universal Approximations', *Neural Networks* **2**, 359–366.
- Isasi, P. & Galván, I. (2004), *Redes neuronales artificiales – un enfoque práctico*, Madrid: Pearson-Prentice Hall.
- Jalil, M. & Misas, M. (2006), Evaluación de pronósticos del tipo de cambio utilizando redes neuronales y funciones de pérdida asimétricas, Technical report, Working Paper 376. Subgerencia de estudios económicos–Banco Central de Colombia.
- Kaastra, I. & Boyd, M. (1996), 'Design a Neural Network for Forecasting Financial and Economic Time Series', *Neurocomputing* **10**, 215–236.
- Maravall, A. & Kaiser, R. (2000), Notes on Time Series Analysis, ARIMA models and Signal Extraction.
*www.bde.es/servicio/software/trabajos.htm
- McLeod, A. I. & Li, W. K. (1983), 'Diagnostic Checking ARMA Time Series Models Using Squared-Residuals Autocorrelation', *Journal of the Time Series Analysis* **4**, 269–273.

- Misas, M., López, E., Arango, C. & Hernández, J. (2003), La demanda de efectivo en Colombia: una caja negra a la luz de las redes neuronales, Technical report, Working Paper 268. Subgerencia de estudios económicos–Banco Central de Colombia.
- Misas, M., López, E. & Borrero, P. (2002), ‘La inflación en Colombia: una aproximación desde redes neuronales’, *Ensayos sobre Política Económica* **41-42**, 143–214.
- Montgomery, D. C. & Johnson, L. A. (1976), *Forecasting and Time Series Analysis*, New York: McGraw-Hill.
- Morettin, P. A. & Toloí, C. M. (2004), *Análise de Séries Temporais*, ABE - Projeto Fisher. São Paulo: Edgard Blücher.
- Narendra, K. & Parthasaranty, K. (1990), ‘Identification and Control of Dynamical Systems Using Neural Networks’, *IEEE Transactions on Neural Networks* **1**, 4–27.
- Ospina, R. M. & Zamprogno, B. (2003), ‘Comparação de Algumas Técnicas de Previsão em Análise de Séries Temporais’, *Revista Colombiana de Estadística* **26(2)**, 129–157.
- R Development Core Team (2005), *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria.
*<http://www.R-project.org>
- Rumelhart, D., Hilton, G. & Williams, R. (1986a), ‘Learning Representations by Backpropagating Errors’, *Nature* **323**, 533–536.
- Rumelhart, D., Hilton, G. & Williams, R. (1986b), *Parallel Distributed Processing*, Cambridge: The MIT Press, chapter Learning representations by backpropagating errors.
- Souza, R. C. & Zandonade, E. (1993), Forecasting Via Neural Networks: Comparative Study, Technical report, Department of Electrical Engineering, Catholic University of Rio de Janeiro.
- Varfis, A. & Versino, C. (1990), *Univariate Economic Time Series Forecasting*, Cambridge University Press.
- Wiegand, A., Huberman, B. & Rumelhart, D. (1990), Predicting the Future: a Connectionist Approach, Technical report, PARC.
- Zhang, G., Patuwo, B. & Hu, Y. (1998), ‘Forecasting with Artificial Neural Networks: The State of Art’, *International Journal of Forecasting* **14**, 35–62.