

Comparación de comportamiento de redes neuronales recurrentes con otros métodos en pronósticos con series de tiempo – caso series hidrológicas.

Carlos Hernán Fajardo-Toro¹, Luis Cobo², Paloma María Teresa Martínez-Sánchez³

chfajardo@universidadean.edu.co, lacobo@universidadean.edu.co, martinezpaloma@unbosque.edu.co

¹ Facultad de administración, Finanzas y ciencias económicas, Universidad Ean, 110221, Bogotá, Colombia.

² Facultad de Ingeniería, Universidad Ean, 110221, Bogotá, Colombia.

² Facultad de Ingeniería, Universidad el Bosque, 110221, Bogotá, Colombia.

Pages: 1022–1034

Resumen: Una de las actividades fundamentales en cualquier organización para poder realizar sus procesos de planificación y presupuestos es la elaboración de pronósticos. Para lograr lo anterior, los responsables de esto deben aplicar técnicas o modelos las cuales pueden ser cualitativos, cuantitativas o una mezcla de ambos. En cuanto a las técnicas cuantitativas, los modelos pueden causales o de series de tiempo, dependiendo de cómo se manejen los datos, manejo que va acorde a como se enfoque el problema. Para ambos casos se utilizan técnicas estadísticas, pero también técnicas de Inteligencia Artificial o adaptativas tales como las redes neuronales o la hibridación de diferentes técnicas. Este artículo se centra en modelos y técnicas para pronósticos de series de tiempo, y para ello presenta una comparación del comportamiento de las redes neuronales recurrentes – RNN tipo Elman contra otras técnicas estadísticas, así como y otras arquitecturas de redes tales como redes perceptrón multicapa o multi-layer perceptron - MLP, FeedForward de una capa, redes de base radial o Radial Basis Function - RBF y una red modular compuesta de redes MLP. Esto se aplica para el cálculo de pronóstico de caudales.

Palabras-clave: Pronósticos, Series de tiempo, Redes neuronales recurrentes, Elman-Jordan, caudales, redes neuronales

Comparison of the behavior of recurrent neural networks against other methods to obtain time series forecasts.

Abstract: One of the fundamental activities of any organization to be able to carry out its planning processes and budgets is the preparation of forecasts. To achieve this, those responsible for the planification must apply techniques or models. These models can be qualitative, quantitative or a mixture of both. As for quantitative

techniques, the models can be causal or for time series analysis; this depending on how the data is handled, according to how the problem is approached. For both cases statistical techniques are used, but also Artificial intelligence techniques such as neural networks or the hybridization of different techniques. This article focuses on models and techniques for time series forecasts, showing a comparison of the behavior of recurrent neural networks, Elman type, against other statistical techniques and different neural network architectures such as Multilayer perceptron - MLP, single layer FeedForward, Radial Basis Function network - RBF and a modular network composed of MLP networks.

Keywords: Forecasts, Time Series, Recurrent Neural Networks, Elman-Jordan, River Flows, neural networks

1. Introducción

Las redes neuronales artificiales – *Artificial Neural Networks* - ANN son algoritmos implementados en forma de programa informático o modelo electrónico, basados en la estructura del cerebro humano. El cerebro aprende fundamentalmente a base de experimentar situaciones. Es un hecho bien conocido que muchos problemas difíciles de resolver para ordenadores, son fácilmente solucionados por pequeños grupos de energía (neuronas) en los cerebros humanos (Haykin, 1994; Corchado et al., 2000). Los modelos del cerebro humano tienen una forma mucho más práctica para solucionar problemas que los proporcionados por máquinas. Incluso los cerebros de pequeños animales son capaces de realizar funciones que ningún ordenador puede resolver, como por ejemplo el reconocimiento de objetos desde diferentes proyecciones. Los ordenadores realizan operaciones mecánicas sin ningún problema, pero tienen problemas para reconocer patrones y más aún para generalizar eventos futuros a partir de acciones pasadas.

Las ANN actuales, intentan únicamente reproducir de forma simplificada los mecanismos más básicos del cerebro animal. No se trata de reproducir el cerebro humano, tan solo se centran en mecanismos de resolución de problemas individuales. El objetivo de las investigaciones teóricas en el campo de las ANN es el de entender la forma en la que los humanos resuelven los problemas, y utilizarla para mejorar las capacidades de los sistemas de Inteligencia Artificial tradicionales. Mientras que el cerebro humano tiene una estructura tridimensional y sus neuronas son capaces de interconectarse de forma incontrolada, las ANN son completamente diferentes, están compuestas por circuitos integrados y tienen un número limitado de capas e interconexiones entre las neuronas de cada nivel. La realidad física restringe el tipo y el ámbito de las ANN.

Las ANN requieren un proceso de entrenamiento previo a su utilización (Corchado et al., 2000). Este proceso comienza con la asignación de valores a los pesos asociados a cada conexión, y la definición de los parámetros de aprendizaje utilizados. En general, los valores de los pesos se eligen de forma aleatoria. Tras realizarse esta operación se lleva a cabo el entrenamiento de la red, lo cual conlleva la adaptación de estos pesos. Existen dos formas de aprendizaje claramente diferenciadas:

- Aprendizaje supervisado (supervised): requiere la presencia de pares de vectores, cada uno de ellos compuesto por los valores de salida de la red correspondientes a los valores de entrada. En este caso, es necesario que el entrenamiento de la red sea controlado por un supervisor o un conjunto de reglas.

- Aprendizaje no supervisado (unsupervised): en el cual se utilizan sólo los vectores de entrada para entrenar la red. Los valores de las salidas son determinados por la red durante el curso del aprendizaje. Se utiliza para construir modelos internos que capturan similitudes en los valores de los vectores de entrada, sin recibir información adicional.

1.1. Predicciones de series temporales

Las ANN poseen la capacidad de proyectar sucesos futuros a partir de experiencias pasadas (Weigend y Gershenfeld, 1994). Existen numerosas aplicaciones en las que se utilizan predicciones; por ejemplo, para ayudar en la definición de prioridades, anticipar futuros acontecimientos, etc. (Bishop, 1995). Predecir el valor de una variable un intervalo hacia delante, utilizando valores previos procedentes de una serie temporal, requiere la estimación de la función autorregresiva que define el proceso. Por tanto, el objetivo de cualquier sistema flexible es el de disminuir el error esperado de la predicción, de tal forma que la complejidad del modelo creado se ajuste a la complejidad de los datos. Por ejemplo, la red MLP y las regresiones B-splines poseen esta capacidad (Canu et al., 1997).

El análisis de series temporales se basa en:

- Predecir: avanzar acertadamente hechos futuros que tendrán lugar en un espacio corto de tiempo.
- Modelar: definir una descripción, que capture las características que definen la evolución de un sistema a largo plazo de forma adecuada.
- Caracterizar: intento de determinar las propiedades fundamentales que definen un sistema, utilizando una pequeña parte del conocimiento o poca información a priori.

Para el problema planteado en este trabajo se han elegido las redes de base radial o RBF (Radial Basis Function), de arquitectura modular, las redes perceptrón multicapa o Multi-layer Perceptron - MLP, la single-Layer feedforward y la Elman-Jordan, como mecanismos para realizar predicciones en series de tiempo; esto como base para el desarrollo de un modelo híbrido en investigaciones posteriores, así como una comparación con los métodos estadísticos tradicionales. Las capacidades de aprendizaje y adaptación a cambios continuos en las variables de entrada, así como su capacidad para manejar entornos no lineales, las convierten en candidatas ideales para implementar una predicción final revisada.

Este documento pretende mostrar como es el comportamiento de las redes neuronales recurrente tipo Elman en predicción de series de tiempo, caso series hidrológicas, comparándolas con otras topologías de redes neuronales y métodos estadísticos tradicionales, mostrando como el comportamiento de estas es bastante bueno con predictoras en este tipo de problemas.

2. Redes recurrentes

Los problemas de predicción constituyen una clase especial de los problemas de aproximación funcional en los que los valores de las variables de salida (predicciones) se determinan a partir de valores de salida obtenidos en instantes anteriores. Para tratar con

esta clase de problemas se introducen unidades de proceso que evolucionan en tiempo discreto, es decir, en el tiempo $t \in \{0,1,2,\dots\}$, de manera que sus salidas son transmitidas como entradas en el tiempo $t+1$. Estas se comportan como un autómata finito.

Sin embargo, el problema está en cómo entrenar dicho autómata para conseguir las secuencias de valores de salida deseados. Las redes neuronales recurrentes constan de una capa de sensores (recogen la información de las entradas), una capa intermedia de unidades ocultas y una capa de unidades de salida; además contienen conexiones sinápticas que van de los sensores de entrada a las unidades ocultas (también pueden ir directamente a las unidades de salida) y de las unidades ocultas a las unidades de salida, y se permiten conexiones sinápticas entre las unidades de proceso de una misma capa sobre todo entre las unidades de la capa oculta.

Se han propuestos varios modelos de redes recurrentes juntos con sus algoritmos de entrenamiento, como los modelos de Elman (1990), Jordan (1986), cuyo algoritmo de aprendizaje es esencialmente el algoritmo de retropropagación.

2.1. Redes Elman

En 1990, Jeffrey L. Elman en su trabajo *Finding Structure in Time* (Elman, 1990) desarrolló una red neuronal basada en el trabajo previo de Jordan (Jordan, 1986). Las redes de Jordan que presentan una topología de capas similar a las redes feedforward, con la inclusión de una capa de contexto que realimenta las señales de salida hacia las capas ocultas. En una red de Jordan, existen retardos en todas las realimentaciones. Adicionalmente, se considera que los pesos de las realimentaciones desde la capa de salida hacia la capa de contexto son unitarios.

Elman modifica esta arquitectura, considerando realimentaciones desde capas ocultas hacia la capa de contexto, sin considerar realimentaciones locales. La capacidad principal, consecuencia de su diseño recursivo, es que puede aprender a detectar y generar patrones temporales; de ahí se desprende su aplicación en procedimiento de señales y otros modelos donde el tiempo juega un papel determinante.

Esta red está constituida por una capa de sensores o unidades de entrada, dos capas intermedias de unidades (una capa de unidades ocultas y otra capa de unidades de contexto) y una capa de unidades de salida. Los sensores de entrada y las unidades de salida recogen la información del entorno, es decir, la suministrada por el conjunto de patrones de entrenamiento, mientras que las demás no. Las unidades de salida reciben la señal suministrada por la salida de las unidades de la capa oculta ponderadas por los correspondientes pesos sinápticos y suelen utilizar como función de transferencia la función identidad.

Sin embargo, la novedad de este modelo la introducen las unidades de contexto que se utilizan para memorizar las salidas de las unidades ocultas en la etapa anterior, de manera que cada unidad de contexto tiene como salida la salida de la unidad oculta correspondiente en la etapa anterior. Por lo tanto, esta red es sólo parcialmente recurrente y cada unidad de proceso oculta recibe como entrada las salidas de las unidades de contexto y de los sensores de entrada ponderadas por sus pesos sinápticos. De esta manera la salida de la red depende no sólo del patrón de entrada actual sino también de los patrones anteriores a través de las unidades de contexto.

Normalmente las funciones de activación de la capa recurrente suelen ser de tipo tangente hiperbólica y las de la capa de salida de tipo lineal. Esta configuración, permite a la red aproximar cualquier función (con un número finito de discontinuidades) con la precisión deseada. La única condición es que la capa oculta debe tener suficientes neuronas. Si la función que se necesita aproximar aumenta de complejidad será necesario incrementar el número de neuronas (Al-Hadithi et al., 2006).

2.2. Redes recurrentes y series de tiempo

Las series de tiempo han sido tratadas tradicionalmente como sistemas lineales, pero desde hace algún tiempo, los sistemas caóticos han venido cobrando importancia. Una señal cualquiera se puede representar a través de sistemas no lineales. Las redes neuronales han sido utilizadas con bastante éxito para solventar este problema, donde las redes recurrentes han mostrado ser bastante eficientes para trabajar con series de tiempo, esto en sus diferentes topologías y características.

Aussen et al., (1994) implementaron redes recurrentes dinámicas en conjunto con métodos para reconocimiento de patrones en la estimación de temperaturas para realizar predicciones meteorológicas. También aplicaron el mismo tipo de redes para la predicción de niveles de visibilidad, variables que en conjunto permiten una mejor programación del telescopio en el observatorio astronómico en Cerro Paranal en Chile.

Ho et al., (2002) contrastan la eficiencia de las redes recurrentes frente a ARIMA en el pronóstico de la cantidad de fallos por unidad de tiempo en una planta en Noruega. More y Deo (2003) comparan la eficiencia de una red recurrente de Jordan con las MLP y los modelos ARIMA en el pronóstico del viento en zonas costeras.

Binner et al., (2006) comparan las redes recurrentes con dos modelos estadísticos utilizados para el cálculo de la inflación en EEUU. Contrasta las redes con modelos de benchmark y Markov Switching Autoregressive models. Jeong et al., (2008) desarrollan un modelo de pronóstico para predecir la dinámica del fitoplancton. Para lograrlo desarrolla una red neural recurrente temporalmente autorregresiva (TARNN, Temporal Autoregressive Recurrent Neural Network) y la contrasta con los modelos SARIMA (Seasonal Auto-Regressive Integrated Moving Average).

Las redes tipo Elman también han demostrado un buen comportamiento con series de tiempo univariadas, y han demostrado ser tan o más eficientes que los modelos estadísticos. Así como también frente a otras topologías como las MLP, las RBF o las SVM (Gençai y Liu, 1997; Recknagel, 2001).

Baker y Richards (1999) demuestran la eficiencia de la red Elman cuando la aplican para calcular el pronóstico del gasto en educación elemental y secundaria en los EEUU. Cho (2002) también lo demuestra al intentar pronosticar el flujo turístico hacia Hong Kong desde otros países. Brunelli et al., (2008) desarrollan un sistema de pronóstico basado en una red Elman para predecir el nivel de dióxido de sulfuro en la región de Priolo en Grecia. Dicho modelo resultó ser más eficiente que los modelos ARIMA.

Dia (2001) lleva a cabo una comparación de 4 redes neuronales para el pronóstico de tráfico de coches en Australia. Hace pruebas con MLP, redes semirecurrentes, redes dinámicas tipo TLRN (Time-Lag Recurrent Network) y redes híbridas. Las redes híbridas

son una combinación de análisis de componentes principales y TLRN. El autor concluye según sus datos, que las redes dinámicas se adaptan mejor a los procesos no lineales que las de tipo estático como las MLP.

Kalaitzakis et al., (2002) utilizan las redes recurrentes para el diseño e implementación de un sistema de pronósticos a corto plazo acerca de la carga de energía del sistema eléctrico de la Isla de Creta. Compara los métodos autorregresivos utilizados con los resultados obtenidos con una MLP, una RBF, una red neuronal recurrente en tiempo real y una red neuronal recurrente autorregresiva para pronósticos de consumo y carga de 24 horas.

Mandal y Prabaharan (2005) construyen una red recurrente que denominan NARX (Nonlinear AutoRegressive models with eXogenous inputs) para el pronóstico de oleajes. Una red NARX consiste en una red MLP que toma como entrada los valores de entrada y salida proporcionales y genera una nueva salida. Calculan los valores de salida y los comparan con los resultados obtenidos en anteriores modelos desarrollados por otros autores obteniendo mejores índices de correlación.

Goh et al., (2006) Abordan el problema del pronóstico del viento, para lo que desarrollan y comparan dos topologías de redes recurrentes. Comparan el desempeño de las redes neuronales concurrentes totalmente conectadas (FCRNN, Fully Connected Recurrent Neural Network) y las redes neuronales recurrentes “entubadas” valoradas complejamente (CPRNN, Complex-valued Pipelined Recurrent Neural Net-works).

Brunelli et al (2007) desarrollan un modelo de pronóstico de concentraciones de compuestos químicos en la atmosfera para el área de Palermo en Italia. Para lograrlo inicialmente implementan dos redes, una red tipo Elman y una MLP, siendo la primera red la que proporciona los pronósticos más acertados.

Cao y Lin (2008) desarrollan una aplicación que permite pronosticar la irradiación del sol utilizando una red neuronal wavelet recurrente diagonal (DRWNN, Diagonal Recurrent Wavelet Neural Network) asistida con una técnica difusa. Una red neuronal wavelet es una ANN integrada con técnicas wavelet como funciones de activación (Cao y Cao, 2006; Benauda et al., 2006). Una red diagonal recurrente es una estructura relativamente simple basada en el algoritmo de retropropagación. No tiene conexiones cruzadas o transversales entre los nodos de la capa oculta o de la capa de salida, reduciendo el número de conexiones y por tanto el tiempo de proceso. Cuando la red se diseña con bases wavelet como funciones de activación, se convierte en una DRWNN (Cao y Lin, 2008).

En esta misma línea, Becerikli y Oysal (2006) diseñan una red denominada TDDNN (Time Delay Dinamic Neural Network) que definen como una red recurrente de tiempos continuos con retroalimentación con valores retardados. Con ella realizan el modelado y pronóstico de series caóticas.

Barbournis y Theocharis (2006) desarrollan su modelo con redes recurrentes para el cálculo de predicciones a largo plazo de la velocidad del viento. Para lograrlo explotan el concepto de las redes localmente recurrente basadas en el algoritmo de predicción recursiva del error. En este tipo de redes la recurrencia es localmente introducida dentro

de cada una de las neuronas de la red, y también se les conoce por el nombre de Local Feedback Multilayer Networks (LF-MLN).

Las redes recurrentes también han presentado una gran eficiencia trabajando con algoritmos evolutivos y genéticos. Pai et al., (2005) se interesa por las máquinas de soporte vectorial recurrente con algoritmos genéticos (RSVMG, Recurrent Support Vector Machine with Genetic Algorithm). Las SVM han sido utilizadas con éxito para resolver problemas no lineales o de series de tiempo, pero no muy frecuentemente en la elaboración de pronósticos (Pai et al., 2005). Los algoritmos genéticos se implementan para determinar los parámetros de la SVM. Los autores realizan una comparación de resultados obtenidos con la RSVMG con los alcanzados por una SVM, una MPL y un modelo autorregresivo, siendo muy superior en su desempeño la RSVMG. Versace et al., (2004), Cai et al., (2007) y Talib et al., (2008) implementan redes recurrentes con algoritmos híbridos evolutivos como algoritmos de aprendizaje, siendo bastantes eficientes a la hora de adaptarse en sistemas caóticos.

3. Diseño experimental y resultados

Los datos brutos utilizados para el presente estudio fueron tomados de la red de monitoreo de la Corporación Autónoma Regional del Valle del Cauca (CVC) que mantiene más de 150 hidrométricos y estaciones de lluvia que realizan la medición y el control de recursos hídricos en el valle y al norte del río Cauca. Bajo un acuerdo con la empresa de energía del pacífico (EPSA), la CVC supervisa el embalse de Salvajina para la producción de energía. En este contexto, uno de los objetivos de CVC es determinar la cantidad de agua que puede ser utilizado para la generación de energía, mientras se restringe el depósito.

Los datos disponibles se reciben y procesan diariamente en la estación hidrométrica de salvajina ubicada en el departamento de Cauca, en Colombia (ver Fig. 1). Esta información representa el registro diario del caudal del río (m³/s), a partir de mediciones de terreno proporcionadas por CVC desde los años 1950 a 2016.

Para generar un conjunto de datos listo para usar para entrenamiento, validación y fines de prueba, era necesario previamente preprocesar y analizar los datos brutos disponibles. El primer paso llevado a cabo fue el tratamiento de los valores equivocados y faltantes. Inicialmente se procedió a depurar los datos inválidos o inexistentes. El primer filtro fue respecto a años con demasiados datos perdidos siguiendo una regla en hidrología que establece que, si falta más de un 15% de los datos, la muestra debe ser anulada. Para las series en las cuales había datos faltantes pero que cumplían con la regla de tener menos del 15%, las series fueron completadas a través de un proceso de ponderación en cual se estima el peso ponderado de ese valor en la serie total, en función del peso del mes frente al total de todo el año.

Para validar el enfoque propuesto, se han probado las redes utilizado el 60% de los datos disponibles para entrenar los modelos, 15% para estimar la mejor configuración parámetros y el restante 25% de los datos para fines de prueba. Esta validación el esquema y la distribución son comúnmente aceptados para probar el rendimiento de diferentes enfoques en el pronóstico de series de tiempo problemas (See et al, 2000).

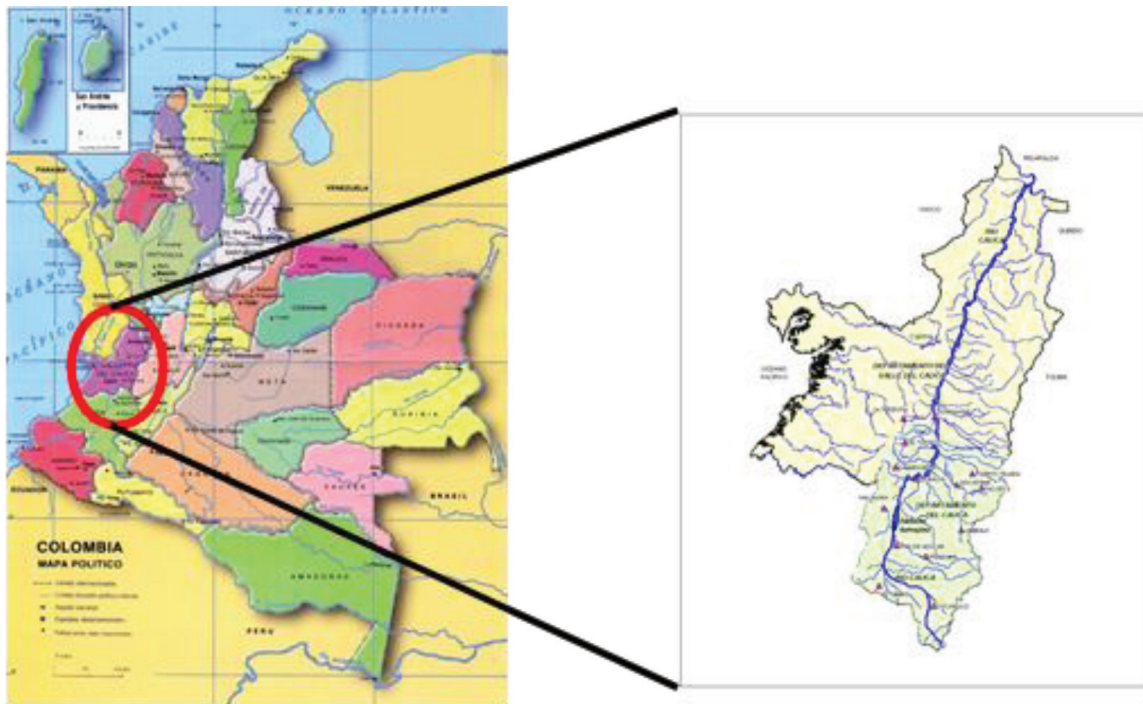


Figura 1 – Localización área de estudio

La topología de la red Elman escogida consta de una capa de entrada con 7 neuronas, que constituyen los días previos al pronóstico a generar, una capa intermedia de 20 neuronas y una capa de salida con una única neurona, correspondiente al caudal a predecir

La red MLP consta también de una capa de entrada de 7 neuronas que constituyen los días previos de pronóstico, y dos capas intermedias de 16 y 8 neuronas respectivamente con una única capa de salida que representa el pronóstico a calcular. La red Modular está compuesta de dos redes MLP que tienen la misma configuración de la red neuronal MLP que se evaluó de forma individual. Para la Single-Layer Feedforward también se tomaron 7 neuronas de entrada y una de salida.

El conjunto de datos de entrada son los 7 días previos al día a pronosticar. El entrenamiento se lleva a cabo con una secuencia de 5000 iteraciones.

Para la Red RBF se realizaron varias simulaciones dando como mejor resultado el que se tomaran 6 días previos con 100 centros y 500 iteraciones.

Una vez realizadas las simulaciones, así como la aplicación de los modelos estadísticos tradicionales, se obtuvieron los siguientes resultados reflejados en las tablas 1 y 2, las cuales presentan los resultados obtenidos utilizando los modelos estadísticos y las redes neuronales. En cuanto a los modelos estadísticos, se dejaron los que dieron mejor resultado, en especial en lo que a ARIMA se refiere. Estos resultados se pueden ver en la Tabla 1. La tabla 2 muestra los resultados de las simulaciones realizadas con las redes neuronales. Como índice de comparación se decidió utilizar la Media absoluta de los errores - MAE.

Modelo	MAE
<i>Naive</i>	18,849
<i>Media constante = 149,671</i>	59,764
<i>Tendencia lineal = 157,513 + -0,00537375 t</i>	55,124
<i>Suavización exponencial simple con alfa = 0,7486</i>	18,11
<i>Suavización exp. De Brown con alfa = 0,4889</i>	18,807
<i>Suavización exp. De Holt con alfa = 0,6327 y beta = 0,0009</i>	19,143
<i>ARIMA(2,0,2) con constante</i>	17,674
<i>ARIMA(1,0,1) con constante</i>	17,916
<i>ARIMA(2,1,2)</i>	17,978
<i>ARIMA(2,1,2) con constante</i>	17,915
<i>ARIMA(2,1,1)</i>	17,026

Tabla 1 – Resumen de errores estadísticos obtenidos para el caudal.

Para evaluar si realmente hay o no diferencias en las medias de los errores entre los métodos, se requiere que hacer pruebas estadísticas sobre la distribución de los errores que demuestren que hay diferencias significativas. Inicialmente realizar una prueba de diferencias de medias es una alternativa válida, pero para aplicarse debe inicialmente cumplir el supuesto de normalidad.

Modelo	MAE
Pronóstico red feedforward	19,992
Pronóstico red Elman	18,345
Pronóstico red RBF	26,911
Pronóstico red Multilayer	58,210
Pronóstico red Modular	22,232

Tabla 2 – Resumen de los índices de errores por redes neuronales para el Caudal

En concreto, para el estudio de las series de error generadas por cada uno de los modelos analizados, se detectó que no tenían un comportamiento normal al realizar las pruebas de Kolmogorov-Smirnov dado el tamaño de muestra. Lo anterior hace que se invalide cualquier test paramétrico que se realice para comparar las muestras. Por este motivo se ha recurrido al contraste de Kruskal-Wallis. Esta prueba representa la alternativa no paramétrica para situaciones como la anteriormente mencionada, centrándose en el análisis de las medianas en lugar del de las medias. La única exigencia que se establece es la aleatoriedad en la extracción de las muestras, no haciendo referencia a ninguna de las otras condiciones adicionales de homocedasticidad y normalidad necesarias para la aplicación cualquier test paramétrico.

Los resultados generados por el contraste de Kruskal-Wallis sobre los datos analizados, muestran, con un nivel de confianza del 99%, que existen diferencias significativas entre los distintos modelos (valor de $p < 0,01$).

Con el fin de profundizar en el análisis de las diferencias significativas encontradas por el contraste de Kruskall-Wallis, se ha realizado la prueba de Mann-Whitney, que permite comparar los sistemas analizados dos a dos, para detectar entre qué modelos existen las diferencias significativas encontradas por la prueba de Kruskall-Wallis.

La Tabla 3 muestra el resumen de la ejecución de la prueba de Mann-Whitney al 99%, para todas las combinaciones posibles de los modelos analizados

Como se podrá observar, las redes Elman y las Feedforward realmente no presentaron diferencias significativas con varias de las técnicas estadísticas, en especial las más robustas como son los modelos ARIMA. Aunque habría que hacer test con más tipos de series de tiempo, se podría decir que el uso de las redes recurrentes resulta de bastante utilidad, dada la dificultad numérica que representan algunos modelos estadísticos, es decir, el grado de experticia en econometría que requiere el decisor para la configuración de los modelos autorregresivos.

4. Conclusiones y trabajo futuro

En este artículo se presentó un comparativo del comportamiento que tienen las redes neuronales, en especial las de tipo recurrente, frente al pronóstico de series de tiempo las cuales presentan un comportamiento estocástico no lineal.

Para lo anterior se tomaron los métodos estadísticos tradicionales y se compararon con 5 redes diferentes, Feedforward, Elman-Jordan, RBF, MLP y una modular compuesta por 2 MLPs internamente, haciendo una comparación de las Medias Absolutas de las Desviaciones – MAE.

	Arima (1,0,1)	Arima (2,0,2)	Arima (2,1,1)	Arima (2,1,2)	Arima (2,1,2)	Naive	M. simple	Brown	Holt	M. Lineal	Elman	FF
Arima(1,0,1)												
Arima(2,0,2)	=											
Arima(2,1,1)	*	*										
Arima(2,1,2)	*	*	=									
Arima(2,1,2)	*	*	=	*								
Naive	*	*	*	*	*							
M. simple	*	*	*	*	*	*						
Brown	*	*	*	*	*	*	*					
Holt	*	=	*	*	*	*	*	*				
M. lineal	=	=	*	*	*	=	*	*	*			
Elman	=	=	=	=	=	=	*	=	*	*		
FeedForward	*	*	=	=	=	*	*	=	=	*	=	

Tabla 2 – Resumen de los índices de errores por redes neuronales para el Caudal

Posterior a esta comparación, y en especial que los errores son bastante similares, se busca realizar pruebas estadísticas que indiquen si existen diferencias representativas

y dado que no tiene un comportamiento normal, se decidió aplicar las pruebas no paramétricas de Kruskal-Wallis y Mann-Whitney las cuales, a través del análisis de medianas, nos pueden indicar si existen diferencias significativas de los errores entre los resultados de las redes neuronales y los métodos estadísticos.

Una vez realizadas las pruebas se logra ver cómo, estadísticamente hablando y con base en las series de tiempo de caudal, no se considera que haya diferencias significativas en la distribución de los errores de los métodos de recurrentes con la mayoría de métodos de mejor precisión estadística.

De lo anterior se puede concluir como este tipo de redes resulta especialmente útil debido a su retropropagación para adaptarse bien al comportamiento de las series de tiempo, que como ya se mencionó, son de naturaleza estocástica no lineal.

Por tanto, se puede considerar una ventaja para los tomadores de decisiones el uso de este tipo de redes en pronósticos dado que, para implementarlas, no se requiere un bagaje matemático amplio como si requiere el conocimiento y definición de modelos de series de tiempo, en especial los autorregresivos.

Como trabajo futuro está el construir modelos más complejos, en los cuales se puedan hacer hibridación de diferentes técnicas de clasificación y pronósticos las cuales contribuyan a disminuir el error o hacer que la distribución de los errores sea más reducida, es decir, con una menor variación con trabajos como los realizados por Toro et al (2013) en la cual se hace uso de sistema de razonamiento basado en casos.

Referencias

- Al-Hadithi B. M., Rueda M. R., Muro J. S. (2006) Interfaz visual para el prototipado rápido de clasificadores de gajos de mandarina basados en redes neuronales, Revista de Ciencia, Tecnología y Medio Ambiente, VOLUMEN IV.
- Aussem, A., Murtagh, F., Sarazin, M. (1994) Dynamical recurrent neural networks and pattern recognition methods for time series prediction: Application to seeing and temperature forecasting in the context of ESO's VLT astronomical weather station. *Vistas in astronomy*, 38(3):357–374.
- Barbounis, T. G. y Theocharis, J. B. Locally recurrent neural networks for long-term wind speed and power prediction. *Neurocomputing*, 69:466–496.
- Becerikli, Y. y Oysal, Y. (2007). Modeling and prediction with a class of time delay dynamic neural networks. *Applied Soft Computing*, 7:1164–1169.
- Binner, J. M., Elger, C. T., Nilsson, B., Tepper, J. A. (2006). Predictable non-linearities in U.S. inflation. *Economics letters*, 93:323–328.
- Bishop, C. R. (1995). *Neural Networks for Pattern Recognition*. Carendon Press, Oxford.
- Brunelli, U., Piazza, V., Pignato, L., Sorbello, F., Vitabile, S. (2007). Two-days ahead prediction of daily maximum concentrations of SO₂, O₃, PM₁₀, NO₂, CO in the urban area of Palermo, Italy. *Atmospheric Environment*, 41:2967–2995.

- Cai, X., Zhang, N., Venayagamoorthy, G. K., Wunsch II, D. C. (2007). Time series prediction with recurrent neural networks trained by a hybrid PSO-EA algorithm. *Neurocomputing*, 70:2342–353.
- Canu, S., Grandvalet, Y., y Ding, X. (1997). Une application des réseaux de neurones pour la prévision à un pas de temps. *Statistiques et méthodes neuronales*. (pp. 120-131), Dunod.
- Cao, J. C. y Cao, S. H. (2006) Study of forecasting solar irradiance using neural networks with preprocessing sample data by wavelet analysis. *Energy*, 31:3435–3445.
- Cao, J. y Lin, X. (2008). Application of the diagonal recurrent wavelet neural network to solar irradiation forecast assisted with fuzzy technique. *Engineering Applications of Artificial Intelligence*, article in press.
- Cao, J. y Lin, X. (2008). Study of hourly daily solar irradiation forecast using diagonal recurrent wavelet neural networks. *Energy Conversion & Management*, 49: 1396–1406.
- Cho, V. (2003) A comparison of three different approaches to tourist arrival forecasting. *Tourism Management*, 24:323–330.
- Corchado, J. M., y Lees, B. (2000). Adaptation of cases for case-based forecasting with neural network support. En Pal, S. K., Dillon, T. S., y Yeung, D. S. (Eds.). *Soft Computing in Case Based Reasoning*. Springer-Verlag: London.
- Dia, H. (2001). an Object-oriented neural network approach to short-term traffic forecasting. *European Journal of operational Research*, 131:253–261.
- Elman, J. L. (1990): Finding structure in time, *Cognitive Science*, 14:179–211
- Gençay R., Liu, T. (1997) Nonlinear modelling and prediction with feedforward and recurrent networks. *Physica D*, 108:119–134.
- Goh, S. L., Chen, M., Popovic, D. H., Aihara, K., Obradovic, D., Mandic, D. P. (2006). Complex-valued forecasting of wind profile. *Renewable Energy*. 31:1733–1750.
- Haykin S. (1994) *Neural Networks, A comprehensive Foundation*. Macmillan College Publishing Company, New York, NY.
- Ho, S. L., Xie, M., Goh, T. N. (2002). A comparative study of neural network and Box-Jenkins ARIMA modeling in time series prediction. *Computers and industrial engineering*, 42:371–375.
- Jordan, M. I. (1986). *Serial order: A parallel distributed processing approach*. Institute for Cognitive Science Report. University of California, San Diego.
- Kalaitzakis, K., Stavrakakis, G. S., Anagnostakis, E. M. (2002). Short-term forecasting based on artificial neural networks parallel implementation. *Electric Power Systems Research*, 63:185–196.
- Mandal, s. y Prabakaran, N. (2006). Ocean wave forecasting using recurrent neural networks. *Ocean Engineering*, 33:1401–1410.

- More, A. y Deo M. C. (2003). Forecasting wind with neural networks. *Marine Structures*, 13:35–49.
- Pai, P. F. y Hong W. C. (2005). Forecasting regional electricity load based on recurrent support vector machines with genetic algorithms. *Electric Power Systems Research*, 74:417–425.
- Recknagel, F. (2001). Applications of machine learning to ecological modelling. *Ecological Modelling*, 146:303–310.
- See L, Openshaw, S, (2000) A hybrid multi-model approach to river level forecasting, *Hydrological Sciences Journal* 45 (4) 523–536.
- Talib, A., Recknagel, F., Cao, H., Van der Molen. D. T. (2008). Forecasting and explanation of algal dynamics in two shallow lakes by recurrent artificial neural network and hybrid evolutionary algorithm. *Mathematics and Computers in Simulation*, 78:424–434.
- Toro CHF, Gómez Meire S, Gálvez JF et al (2013) A hybrid artificial intelligence model for river flow forecasting. *Appl Soft Computing* 13:3449–3458
- Versace, M., Bhatt, R., Hinds, O., Shiffer, M. (2004). Predicting the exchange traded fund DIA with combination of genetic algorithms and neural networks. *Expert systems with Applications*, 27:417–425.
- Weigend, A. S., y Gershenfeld, N. A. (1994). *Time series prediction: Forecasting the future and understanding the past*. Santa Fe institute studies in the sciences of complexity, Addison-Wesley Publishing Company.