

Una aproximación a la regularización de redes cascada-correlación para la predicción de series de tiempo

Fernan A. Villa †
Juan D. Velásquez ‡
Reinaldo C. Souza *

† Grupo de Computación Aplicada
Universidad Nacional de Colombia – Sede Medellín, Colombia
fernnavilla@gmail.com

‡ Grupo de Computación Aplicada
Universidad Nacional de Colombia – Sede Medellín, Colombia
jdvelasq@unal.edu.co
www.docentes.unal.edu.co/jdvelasq

* Departamento de Engenharia Elétrica
Pontificia Universidade Católica (PUC-Rio) Rio de Janeiro – RJ, Brasil
reinaldo@ele.puc-rio.br

Abstract

Forecasting of time series using artificial neural networks is an important research topic due to the practical implications in fields as economics, finance and social sciences. Cascade-correlation neural networks seem to have better abilities for capturing nonlinear dynamics in relation to the other classical architectures as multilayer perceptrons. However, cascade-correlation network, as other models, may over fit the data. In this paper, we compare the ability of cascade-correlation networks trained using regularization techniques for forecasting a benchmark time series, and we show that regularization techniques allows us to find models with better generalization and forecasting ability.

Resumen

La predicción de series de tiempo usando redes neuronales artificiales es un tópico importante de investigación debido a sus implicaciones prácticas en campos como la economía, las finanzas y las ciencias sociales. Las redes neuronales cascada-correlación parecen tener mejores habilidades para capturar dinámicas no lineales en comparación con otras arquitecturas clásicas tales como los perceptrones multicapas. Sin embargo, la red cascada-correlación, como otros modelos, podría sobreajustar los datos. En este artículo, se compara la habilidad de las redes cascada-correlación, entrenadas usando técnicas de regularización, para pronosticar una serie benchmark, y se muestra que las técnicas de regularización permiten encontrar modelos con mejor generalización y habilidad de pronóstico.

Keywords: cascade-correlation neural network, regularization, weight decay, weight elimination, time series, prediction

Title: An approximation to regularization of cascade-correlation neural networks for time series prediction.

1 Introducción

Los perceptrones multicapa (MLP, por su sigla en inglés) son aproximadores universales de funciones que estén definidas en un dominio compacto [Hornik, Stinchcombe y White, 1989; Cybenko, 1989; Funahashi, 1989]. No obstante, el proceso de especificación de un MLP es difícil debido a la gran cantidad de pasos metodológicos que requiere, a los criterios subjetivos en cuanto a cómo abordar cada paso, y a que los resultados obtenidos en cada etapa son críticos. Uno de los aspectos que dificultan el proceso de especificación es la falta de identificabilidad estadística del modelo. Las consideraciones sobre éste tema son el punto inicial para plantear modificaciones sobre la especificación de los MLP, tal que se obtengan nuevas configuraciones que puedan modelar problemas de una forma más objetiva, elegante y simple; y que permitan obtener mejores resultados en comparación con otros modelos. Desde este punto de vista, la red neuronal artificial conocida como Cascada-Correlación (CC) propuesta por Fahlman y Lebiere [1990] presenta ventajas conceptuales interesantes en relación al problema de identificabilidad y capacidad de generalización de los MLP.

La elección del tamaño óptimo de una red neuronal es un paso crítico al modelar cualquier problema: si se elige una red de tamaño relativamente pequeño no será capaz de generalizar con precisión los datos y, por tanto, no será capaz de aprender las características más importantes inmersas en los datos. En consecuencia, es necesario aumentar el tamaño de la red. Mientras que una red de un tamaño innecesariamente grande tiende a aprender no sólo las características de los datos dados, sino también el ruido y la idiosincrasia de los mismos. En aquel momento, la red incurre en sobre-ajuste y su tamaño debe ser reducido. El sobre-ajuste está relacionado con la saturación de las neuronas y se evidencia cuando se produce un error de entrenamiento muy pequeño y un error de validación muy alto. Este problema es controlado en los MLP principalmente mediante dos enfoques de regularización: el crecimiento de la red (Network Growing) y el podado o reducción de la red (Network Pruning) [Palit y Pppovic, 2005]. El enfoque de crecimiento de la red consiste en comenzar con una red de tamaño mínimo y agregar sucesivamente nuevas neuronas hasta lograr un rendimiento deseado. En la reducción de la red se comienza con una red relativamente grande y sucesivamente se van eliminando o anulando neuronas de acuerdo a un criterio definido, hasta que el desempeño de la red se degenere; una de las más importantes críticas a este método es que no se sabe si la red inicial es suficientemente grande como para que tenga neuronas innecesarias.

El enfoque de reducción de la red se usa preferiblemente cuando se desea diseñar redes que posean una gran capacidad de generalización, por ejemplo, para problemas como la predicción de series temporales o la clasificación de patrones, entre otros (Palit y Pppovic, 2005). En éste enfoque se tienen principalmente las estrategias de descomposición de pesos (Weight Decay) propuesta por Hinton [1989] y eliminación de pesos (Weight Elimination) por Weigent, Rumelhart y Huberman [1991]; las cuales han sido ampliamente utilizadas para regularizar los MLP. Sin embargo, no se han considerado el uso de dichas estrategias para el entrenamiento de las redes CC, aunque es de esperarse que podrían obtenerse modelos con una mejor capacidad de generalización.

Este artículo tiene dos objetivos: primero, presentar una modificación de CC para incorporar las estrategias de regularización mencionadas anteriormente: descomposición y eliminación de pesos; y segundo, analizar empíricamente las ganancias al pronosticar una serie de tiempo mediante modelos de CC regularizados.

Para cumplir con los objetivos propuestos, el resto del artículo está organizado así: en la siguiente sección se discuten las técnicas de regularización; posteriormente, se hace una introducción a las redes CC y se mencionan las estrategias de regularización a integrar; luego, se pronostica una serie de tiempo para comprobar la efectividad de las modificaciones propuestas y finalmente se concluye.

2 Las técnicas de regularización

La solución de problemas complejos, como la predicción de series temporales o la clasificación de patrones, exige el uso de redes gran tamaño y altamente estructuradas [Haykin, 1999]. Sin embargo, las redes de gran tamaño son propensas a aprender las particularidades o ruido presente en los datos de entrenamiento y a incurrir en el problema bien conocido del sobre-ajuste. Además, el procesamiento computacional requerido sea mayor respecto a otras redes de menor tamaño. Una manera de sortear estas dificultades es buscar la minimización del tamaño de la red mientras se mantiene su buen rendimiento; esto se puede lograr a través del enfoque de regularización basado en la reducción de red usando técnicas de poda [Haykin, 1999; Palit y Popovic, 2005]. Así, se puede llegar a tener una red con un tamaño óptimo, menos propensa a aprender el ruido en los datos de entrenamiento y a incurrir en el sobre-ajuste. Consecuentemente, una red de tamaño óptimo puede generalizar con mayor precisión en un tiempo computacional menor que una red de mayor tamaño.

Por otro lado, si se elige una red de tamaño relativamente pequeño, esta no será capaz de generalizar con precisión los datos y, por tanto, no será capaz de aprender sus características más importantes. En consecuencia, es necesario aumentar el tamaño de la red y es recomendable seguir el enfoque de regularización: crecimiento de red (Network Growing) [Haykin, 1999; Palit y Popovic, 2005].

En el enfoque de crecimiento de la red se comienza con una red pequeña; luego, se agregan secuencialmente nuevas neuronas o capas ocultas hasta que la red logre un rendimiento adecuado. En las técnicas de reducción, se comienza con una red de gran tamaño y seguidamente se eliminan secuencialmente conexiones de manera selectiva y ordenada; la eliminación se puede lograr a través de una de dos estrategias: la descomposición de pesos (Weight Decay) propuesta por Hinton [1989] y eliminación de pesos (Weight Elimination) propuesta por Weigent et al. [1991]; ambas ampliamente utilizadas para regularizar los MLP.

Entonces, las estrategias de regularización tienen como objetivo realizar un intercambio apropiado entre la fiabilidad de los datos de entrenamiento y las bondades del modelo. En procedimientos de aprendizaje supervisado, el intercambio se realiza a través de la minimización el riesgo total [Haykin, 1999], dado por la expresión:

$$R(w) = \xi_s(W) + \lambda \xi_c(w) \quad (1)$$

Donde $\xi_s(W)$ es la medida estándar de rendimiento, depende del modelo de la red y de los datos de entrada, en aprendizaje backpropagation es conocido como la media del cuadrático; λ es el parámetro de regularización; $\xi_c(w)$ es la penalización compleja, para una red en general, está dado por una integral de suavizado de orden k , así:

$$\xi_c(w, k) = \frac{1}{2} \int \left\| \frac{\partial^k}{\partial x^k} F(x, m) \right\|^2 \mu(x) dx \quad (2)$$

donde $F(x, m)$ es el mapeo de entrada-salida realizado por el modelo, $\mu(x)$ es alguna función de ponderación que determina la región del espacio de entrada sobre la cual la función $F(x, m)$ es requerida para ser suavizada.

2.1 La descomposición de pesos (Weight Decay)

El procedimiento de descomposición de pesos propuesto por Hinton [1989], es un método de regularización complejo; opera sobre algunos pesos sinápticos de la red forzándolos a tomar valores cercanos a cero y permitiendo a otros conservar valores relativamente altos. Esta discriminación permite agrupar los pesos de la red en: pesos que tienen poca o ninguna influencia sobre el modelo, llamados pesos de exceso; y pesos que tienen influencia sobre el modelo. Para ésta técnica el procedimiento la penalización de complejidad es definido como:

$$\xi_c(w) = \|w\|^2 = \sum_{i \in \zeta_{total}} w_i^2 \quad (3)$$

donde ζ_{total} son los pesos sinápticos en la red. El tratamiento de los pesos de la red CC es similar al de los MLP; todos los pesos son tratados igual, es decir, se supone que la distribución de los pesos en el espacio estará centrada en el origen.

2.2 La eliminación de pesos (Weight Elimination)

Este método de regularización descrito por Weigend et al. [1991] define la penalización de complejidad como:

$$\xi_c(w) = \|w\|^2 = \sum_{i \in \zeta_{total}} w_i^2 \quad (4)$$

donde w_i es el peso de alguna sinapsis i en la red; w_0 es un parámetro predefinido; y ζ_{total} son todas las conexiones sinápticas en la red. El término w_i/w_0 hace que la penalización compleja tenga un comportamiento simétrico. Además, cuando $|w_i| \ll w_0$, $\xi_c(w)$ tiende a cero, es decir, para el aprendizaje el peso sináptico i es poco fiable, por consiguiente puede ser eliminado de la red. Mientras que cuando $|w_i| \gg w_0$, $\xi_c(w)$ tiende a uno, entonces el peso w_i es importante para el proceso de aprendizaje. En conclusión, éste método busca los pesos que tienen una influencia significativa sobre la red, y descarta los demás.

3 Redes cascada-correlación

La red neuronal artificial conocida como Cascada Correlación (CC) propuesta por Fahlman y Lebiere [1990], está diseñada siguiendo el esquema de crecimiento de red, se comienza con una red mínima sin capas ocultas, es decir, con sólo algunas entradas y uno o más nodos de salida. Las neuronas ocultas son agregadas una a una en la red, obteniendo de esta manera una estructura multicapa, que permite aplicar las técnicas de

regularización utilizadas para perceptrones multicapa. En la Figura 1 se presenta pictóricamente el esquema de una red CC.

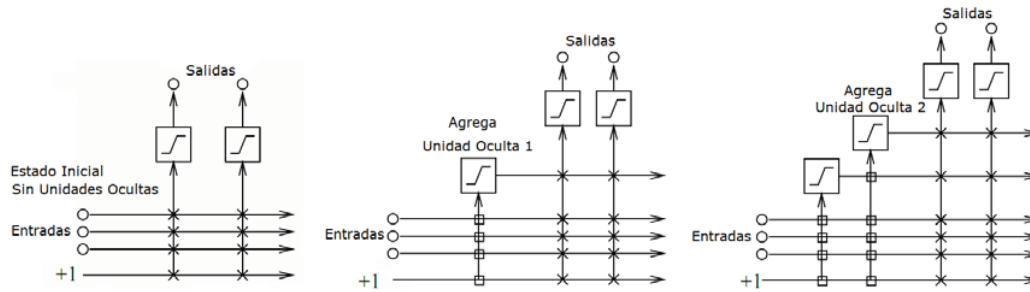


Figura 1: Esquema de una red Cascada-Correlación, por Fahlman & Lebiere, 1991.

En el proceso de adición de neuronas ocultas a la red, cada nueva neurona recibe una conexión sináptica de cada una de las neuronas de entrada y también de las neuronas ocultas que la preceden. Luego de agregar la nueva neurona oculta, los pesos sinápticos de su entrada son congelados, mientras que los pesos de su salida son entrenados repetidamente. Este proceso es continuo hasta que se alcanza un rendimiento satisfactorio.

Consecuentemente, una red CC podría realizar la regresión de funciones no lineales con una precisión superior al de un MLP tradicional. Esto (el problema general de regresión) ya ha sido abordado en la literatura; pero, el problema del modelado y la predicción de series temporales es más complejo que el problema de regresión, porque se debe tener en cuenta el orden de los datos así como las nuevas propiedades estadísticas que este ordenamiento induce sobre la información. Además, no se han considerado estrategias de regularización para las redes CC.

4 Incorporación de las estrategias de regularización

Para la regularización de las redes CC se sigue el enfoque de reducción de la red, puesto que se usa preferiblemente cuando se desea diseñar redes que posean una gran capacidad de generalización [Palit y Popovic, 2005]. Bajo este enfoque se tienen principalmente las estrategias de Descomposición de Pesos y Eliminación de Pesos.

Para el caso de la regularización de CC se consideran las expresiones: reemplazando la ecuación (3) en (1), se puede incorporar la estrategia de descomposición de pesos:

$$R(w) = \xi_s(W) + \lambda \left(\sum_{i \in \zeta_{total}} w_i^2 \right) \tag{5}$$

Mientras que reemplazando (4) en (1), se puede incorporar la estrategia de eliminación de pesos:

$$R(w) = \xi_s(W) + \lambda \left(\sum_{i \in \zeta_{total}} \frac{(w_i / w_0)^2}{1 + (w_i / w_0)^2} \right) \tag{6}$$

5 Caso de estudio

En ésta sección se presenta una comparación entre una red cascada correlación sin regularizar y varias redes CC regularizadas con los esquemas de eliminación y descomposición al pronosticar la serie de tiempo “Pasajeros de una Aerolínea” de Box y Jenkins [1976]. Esta serie ha sido estudiada en la literatura por Faraway y Chatfield [1998] utilizando un MLP, por Ghiassi, Saidane y Zimbra [2005] mediante DAN2 y por Ortiz, Villa y Velásquez [2007]. La serie posee un comportamiento no lineal, como se puede apreciar en las Figura 2, y corresponde al registro del número total de pasajeros transportados por mes por una aerolínea, desde enero de 1949 hasta diciembre de 1960. En la Figura 2 se muestran los valores reales y los pronosticados de la serie.

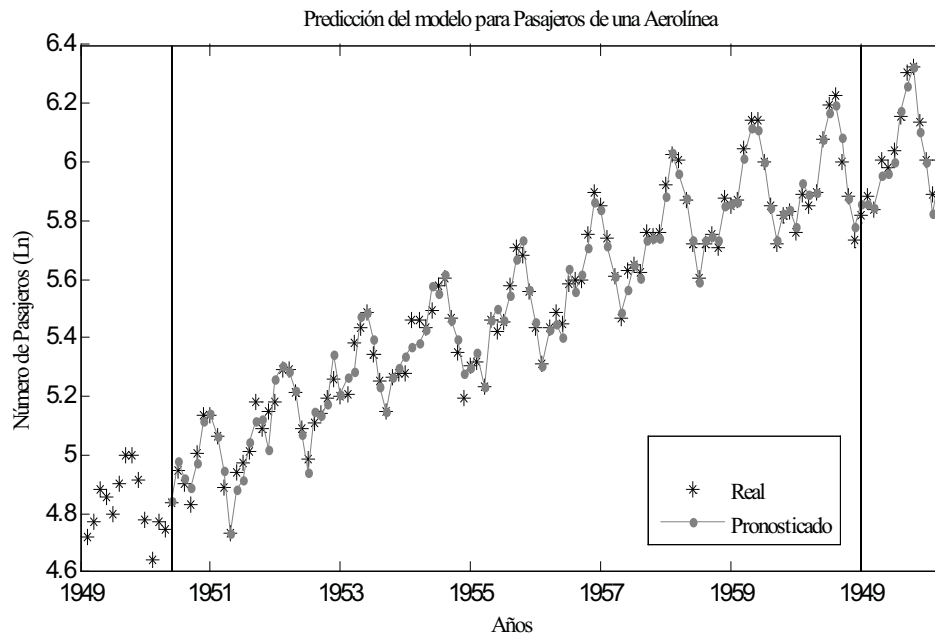


Figura 1: Valores real y pronosticado para la serie de pasajeros de una aerolínea.

Para comparar la habilidad de las redes CC sin regularizar y regularizadas, se calcula la sumatoria del error medio cuadrático (SSE) de entrenamiento y validación, al pronosticar la serie de tiempo con 17 modelos (Tabla 3) de redes CC: sin regularizar; y regularizadas con descomposición y eliminación de pesos, sus parámetros se indican en la Tablas 1 y 2, respectivamente. Además, los datos de la serie se transformaron utilizando la función logaritmo natural (base - e); para el pronóstico, se usaron los primeros 120 datos para entrenamiento y los 12 últimos para validación. Los algoritmos se implementaron en Matlab®.

En las Tablas 3 y 4 se resumen los resultados de entrenamiento y validación, respectivamente; al regularizar mediante descomposición de pesos (CC-D_i). Mientras que en las Tablas 5 y 6 se presentan los resultados al pronosticar con redes CC regularizadas con eliminación de pesos (CC-E_j). Para las Tablas 3–6, la columna CC indica que el pronóstico se realizó sin ninguna estrategia de regularización.

Tabla 1: Parámetros de regularización para el esquema de regularización de descomposición de pesos.

Descomposición de Pesos				
Parámetro	CC-D ₁	CC-D ₂	CC-D ₃	CC-D ₄
λ	0.001	0.010	0.050	0.100

Tabla 2: Parámetros de regularización para el esquema de regularización de eliminación de pesos.

Eliminación de Pesos								
Parámetros	CC-E ₁	CC-E ₂	CC-E ₃	CC-E ₄	CC-E ₅	CC-E ₆	CC-E ₇	CC-E ₈
λ	0.001	0.010	0.050	0.100	0.001	0.010	0.050	0.100
$w0$	10	10	10	10	100	100	100	100

Tabla 3: Sumatoria del error medio cuadrático en entrenamiento para diferentes modelos regularizados con la estrategia de descomposición de pesos, pronosticando la serie del caso de estudio.

Entrenamiento							
Modelo	Rezagos	Neuronas	CC	CC-D ₁	CC-D ₂	CC-D ₃	CC-D ₄
1	1, 2, 13	4	0.826	0.612	1.256	1.420	1.459
2	1, 4, 8, 12	3	0.228	0.337	0.467	0.844	1.031
3	1, 4, 8, 12, 13	4	0.106	0.223	0.513	0.849	0.983
4	1, 4, 8, 10, 12, 13	3	0.171	0.221	0.491	0.816	0.957
5	1 – 4	2	1.171	1.123	1.487	2.062	2.277
6	1 – 13	2	0.145	0.214	0.451	0.821	1.057
7	1 – 13	4	0.174	0.214	0.451	0.821	1.057
8	1, 12	2	0.301	0.343	0.391	0.435	0.454
9	1, 12	4	0.286	0.343	0.391	0.435	0.454
10	1, 12	10	0.242	0.343	0.391	0.435	0.454
11	1, 2, 12	2	0.334	0.335	0.457	0.690	0.768
12	1, 2, 12	4	0.255	0.335	0.457	0.690	0.768
13	1, 2, 12, 13	2	0.185	0.223	0.502	0.783	0.863
14	1, 2, 12, 13	4	0.184	0.223	0.502	0.783	0.863
15	1, 12, 13	1	0.183	0.223	0.473	0.644	0.684
16	1, 12, 13	2	0.186	0.223	0.473	0.644	0.684
17	1, 12, 13	4	0.154	0.223	0.473	0.644	0.684

Tabla 4: Sumatoria del error medio cuadrático en validación para los modelos de la Tabla 3 regularizados con la estrategia de descomposición de pesos, pronosticando la serie del caso de estudio.

Modelo	Rezagos	Neuronas	Validación				
			CC	CC-D1	CC-D2	CC-D3	CC-D4
1	1, 2, 13	4	0.196	0.139	0.148	0.164	0.169
2	1, 4, 8, 12	3	0.036	0.022	0.031	0.079	0.106
3	1, 4, 8, 12, 13	4	0.020	0.014	0.036	0.078	0.096
4	1, 4, 8, 10, 12, 13	3	0.014	0.013	0.031	0.067	0.085
5	1-4	2	0.140	0.162	0.174	0.239	0.266
6	1-13	2	0.059	0.016	0.028	0.068	0.102
7	1-13	4	0.013	0.016	0.028	0.068	0.102
8	1, 12	2	0.033	0.020	0.028	0.035	0.038
9	1, 12	4	0.019	0.020	0.028	0.035	0.038
10	1, 12	10	0.046	0.022	0.028	0.035	0.038
11	1, 2, 12	2	0.023	0.022	0.032	0.062	0.073
12	1, 2, 12	4	0.039	0.022	0.032	0.062	0.073
13	1, 2, 12, 13	2	0.012	0.014	0.036	0.071	0.082
14	1, 2, 12, 13	4	0.012	0.014	0.036	0.071	0.082
15	1, 12, 13	1	0.011	0.014	0.036	0.057	0.062
16	1, 12, 13	2	0.011	0.014	0.036	0.057	0.062
17	1, 12, 13	4	0.010	0.014	0.036	0.057	0.062

Los resultados presentados en la Tabla 3 indican que al hacer $\lambda=0.001$ (columna CC-D₁), es indiferente utilizar el modelo 6 ó 7 para entrenamiento, dado que logran mismos errores. Asimismo, es indiferente usar los modelos: 8, 9 ó 10; 11 ó 12; y 13, 14, 15, 16 ó 17; son claramente 4 grupos de modelos. Al aumentar λ a 0.01 (CC-D₂), es indiferente utilizar en entrenamiento: 6 ó 7; 8, 9 ó 10; 11 ó 12; 13 ó 14; 15, 16 ó 17; son 5 grupos. Haciendo $\lambda = 0.05$ (CC-D₃) se distinguen los mismos grupos de D₂ pero con un error mayor, igualmente cuando se aumenta λ a 0.1 (CC-D₄) también aumenta el error. Además, en la validación (Tabla 4), similar al entrenamiento en varios modelos el error obtenido fue igual; en CC-D₁ se tienen 4 grupos de modelos con el mismo error, en CC-D₂ 3 grupos, y en CC-D₃ y CC-D₄ 4 grupos. Tanto en entrenamiento como en validación, se observa que la descomposición de pesos logra que los errores varíen menos entre modelos, esto posibilita agruparlos y que sea indiferente utilizar cualquier modelo de un grupo específico.

Además, en entrenamiento y validación, los modelos 1-5, donde varían la cantidad de neuronas y los rezagos, los errores con CC-D₁ son cercanos a los obtenidos con CC e incluso algunos son menores (en entrenamiento los modelos 1 y 5; y en validación 1, 2, 3 y 4). Sin embargo, cuando se aumenta λ (se hace que el término de regularización tenga más importancia en la red) los errores aumentan, tal es el caso de las columnas CC-D₂, CC-D₃, CC-D₄. En los modelos 6 y 7, se mantienen fijos los rezagos, y al aumentar las neuronas ocultas el error de entrenamiento no varía, pero si cambia en redes CC sin regularizar. Similarmente, en los modelos: 8, 9 y 10, al aumentar las neuronas ocultas, primero en dos unidades, y luego en seis, los errores no cambian; 13 y 14 las neuronas se incrementan en dos unidades y los errores permanecen estables; y 15, 16 y 17, de 15 a 16 se aumenta una neurona, luego de 16 a 17 dos unidades y ocurre lo mismo.

Tabla 5: Sumatoria del error medio cuadrático de entrenamiento para los modelos de la Tabla 3 regularizados con la estrategia de eliminación de pesos, pronosticando la serie del caso de estudio.

Modelo	Entrenamiento								
	CC	CC-E ₁	CC-E ₂	CC-E ₃	CC-E ₄	CC-E ₅	CC-E ₆	CC-E ₇	CC-E ₈
1	0.826	0.732	0.729	0.874	0.642	1.002	0.803	1.030	0.736
2	0.228	0.227	0.222	0.328	0.210	0.332	0.195	0.337	0.231
3	0.106	0.103	0.101	0.169	0.107	0.197	0.109	0.222	0.108
4	0.171	0.118	0.086	0.174	0.111	0.194	0.114	0.220	0.119
5	1.171	0.870	0.520	0.936	0.934	1.060	0.889	0.846	0.882
6	0.145	0.089	0.081	0.150	0.084	0.188	0.082	0.213	0.094
7	0.174	0.116	0.115	0.155	0.117	0.188	0.123	0.213	0.119
8	0.301	0.305	0.267	0.315	0.291	0.341	0.293	0.343	0.300
9	0.286	0.276	0.239	0.316	0.284	0.341	0.276	0.343	0.280
10	0.242	0.221	0.211	0.296	0.229	0.341	0.216	0.343	0.215
11	0.334	0.244	0.222	0.308	0.224	0.331	0.244	0.335	0.244
12	0.255	0.223	0.207	0.305	0.230	0.331	0.191	0.335	0.201
13	0.185	0.162	0.147	0.174	0.137	0.197	0.154	0.223	0.159
14	0.184	0.136	0.131	0.166	0.122	0.197	0.139	0.223	0.161
15	0.183	0.181	0.171	0.184	0.178	0.198	0.176	0.223	0.181
16	0.186	0.161	0.138	0.179	0.129	0.198	0.170	0.223	0.163
17	0.154	0.143	0.119	0.168	0.125	0.198	0.137	0.223	0.143

Tabla 6: Sumatoria del error medio cuadrático de validación para los modelos de la Tabla 3 regularizados con la estrategia de eliminación de pesos pronosticando la serie del caso de estudio.

Modelo	Validación								
	CC	CC-E ₁	CC-E ₂	CC-E ₃	CC-E ₄	CC-E ₅	CC-E ₆	CC-E ₇	CC-E ₈
1	0,196	0,141	0,179	0,189	0,526	0,145	0,236	0,139	0,368
2	0,036	0,038	0,035	0,025	0,052	0,023	0,045	0,022	0,038
3	0,020	0,024	0,018	0,017	0,026	0,014	0,017	0,014	0,027
4	0,014	0,042	0,024	0,017	0,035	0,015	0,035	0,015	0,041
5	0,140	0,156	0,101	0,164	0,158	0,140	0,148	0,140	0,152
6	0,059	0,014	0,014	0,016	0,011	0,015	0,013	0,016	0,026
7	0,013	0,023	0,018	0,017	0,012	0,015	0,013	0,016	0,026
8	0,033	0,026	0,030	0,022	0,019	0,020	0,030	0,020	0,051
9	0,019	0,037	0,037	0,023	0,029	0,020	0,030	0,020	0,051
10	0,046	0,050	0,050	0,028	0,059	0,020	0,030	0,020	0,051
11	0,023	0,037	0,040	0,027	0,036	0,023	0,046	0,022	0,037
12	0,039	0,049	0,075	0,028	0,042	0,023	0,046	0,022	0,037
13	0,012	0,019	0,019	0,016	0,020	0,014	0,026	0,014	0,016
14	0,012	0,024	0,022	0,018	0,022	0,014	0,026	0,014	0,016
15	0,011	0,015	0,015	0,013	0,014	0,013	0,023	0,014	0,015
16	0,011	0,025	0,016	0,014	0,016	0,013	0,023	0,014	0,015
17	0,010	0,027	0,029	0,016	0,026	0,013	0,023	0,014	0,015

Consecuentemente, los resultados experimentales (entrenamiento y validación) al pronosticar la serie con redes CC regularizadas mediante descomposición de pesos muestran que: se logra un error estable a pesar de que se aumente la cantidad de neuronas ocultas en un modelo de red CC; con λ relativamente pequeño ($\lambda=0.001$) se pueden lograr errores menores que los obtenidos con redes CC sin regularizar; al aumentar λ los errores continúan siendo estables, pero aumentan.

En la Tabla 5 se presentan los resultados experimentales de entrenamiento al pronosticar la serie con los 17 modelos de la Tabla 3 regularizados con la estrategia de eliminación de pesos, variando el parámetro λ y w_0 como se indica en la Tabla 2. Los resultados revelan que dejando $w_0=10$ fijo el error de entrenamiento de la red CC no regularizada se disminuye al hacer $\lambda=0.001$ (columna CC-E₁), se reduce aún más cuando se aumenta λ a 0.01 (CC-E₂) en todos los modelos. Sin embargo, el error incrementa cuando se aumenta λ a 0.05 (CC-E₃), pero al incrementar λ a 0.1 (CC-E₄), el error disminuye respecto a CC-E₃, es decir, CC-E₄ < CC-E₃. Luego, cuando w_0 es aumentado a 100 y se mantienen fijo, se nota que algunos modelos tienden a un error específico aunque se aumenten el número de neuronas, si $\lambda=0.001$ (CC-E₅) los modelos: 6 y 7 tienen un error de 0.188; 8, 9 y 10 de 0.341; 11 y 12 de 0.331; 13 y 14 de 0.197; y 15, 16 y 17 de 0.198. Igualmente, cuando $\lambda=0.05$ (CC-E₇) los modelos tienden al mismo error, pero mayor que el logrado con $\lambda=0.001$. Además, con $\lambda=0.01$ (CC-E₆) y $\lambda=0.1$ (CC-E₈) los errores obtenidos son menores que los logrados con redes CC sin regularizar.

Los resultados en validación (Tabla 6) al pronosticar con el esquema de eliminación de pesos muestran que cuando $w_0=10$ los errores de las columnas CC-E₁, CC-E₂, CC-E₃, CC-E₄ son relativamente cercanos a los obtenidos con redes CC sin regularizar, e incluso algunos son menores; sin embargo, al aumentar el número de neuronas ocultas el error aumenta. Mientras que si w_0 se aumenta a 100, los errores de los modelos tienden a un error, aunque se aumente el número de neuronas, y en algunos casos es menor al de las redes CC sin regularizar.

Entonces, los resultados experimentales (entrenamiento y validación) al pronosticar la serie con redes CC regularizadas mediante eliminación de pesos muestran que: cuando $w_0=100$ se logra un error estable a pesar de que se aumente la cantidad de neuronas ocultas en un modelo de red CC; y con diferentes combinaciones de λ y w_0 , e.g. $\lambda=0.01$ y $w_0=100$, se pueden lograr errores menores que los obtenidos con redes CC sin regularizar.

6 Conclusiones

Los resultados experimentales al realizar el pronóstico de la serie del caso de estudio con redes CC regularizadas mediante descomposición de pesos muestran, tanto en entrenamiento como en validación, que: aunque se aumente el número de neuronas en el modelo de CC, éste sigue tendiendo al mismo error; relativamente pequeño ($\lambda=0.001$) se pueden lograr errores menores que los obtenidos con redes CC sin regularizar; al aumentar λ los errores continúan siendo estables, pero aumentan.

Mientras que al pronosticar la serie usando como estrategia de regularización la eliminación de pesos los resultados en entrenamiento y validación, muestran que: cuando $w_0=100$ se logra un error estable a pesar de que se aumente la cantidad de neuronas ocultas en un modelo de red CC; con diferentes combinaciones de λ y w_0 , e.g. $\lambda=0.01$ y $w_0=100$, se pueden lograr errores menores que los obtenidos con redes CC sin regularizar.

Consecuentemente, es favorable incorporar estrategias de regularización en el diseño de las redes CC; además, tal incorporación aporta al problema del pronóstico de series de tiempo.

7 Reconocimientos

Este artículo se realizó en el marco del proyecto de investigación: “MODELADO Y PREDICCIÓN DE SERIES TEMPORALES NO LINEALES USANDO REDES CASCADA-CORRELACION”, financiado por la DIME – Universidad Nacional de Colombia (Medellín).

8 Referencias

- Box, G.; Jenkins, G. [1976] Time series analysis, forecasting and control. Holden-Day.
- Cybenko, G. [1989] Approximation by superpositions of a sigmoidal function. *Mathematics of Control: Signals and Systems*, Vol. 2, pp. 202–314.
- Fahlman, S. E.; Lebiere C. [1990] The Cascade-Correlation learning architecture. *Advances in Neural Information Processing Systems*, Vol. 2, pp. 524–532.
- Faraway, J.; Chatfield, C. [1998] Time series forecasting with neural networks: A comparative study using the airline data. *Applied Statistic*, Vol. 47, Nro. 2, pp. 231–250.
- Funahashi, K. [1989] On the approximate realization of continuous mappings by neural networks. *Neural Networks*, Vol. 2, pp. 183–192.
- Ghiassi, M.; Saidane, H.; Zimbra, D [2005] A dynamic artificial neural network model for forecasting time series events. *International Journal of Forecasting*, Vol. 21, No. 2, pp. 341–362.
- Haykin, S. [1999] *Neural Networks: A Comprehensive Foundation*. Prentice Hall.
- Hinton, G. E. [1989] Connectionist learning procedures. *Artificial Intelligence*, No. 40, pp. 185–243.
- Hornik, K.; Stinchcombe, M.; White, H. [1989] Multilayer feedforward networks are universal approximators. *Neural Networks*, Vol. 2, No. 5, pp. 359–366.
- Ortiz, D. M.; Villa, F. A.; Velásquez, J. D. [2007] Una Comparación entre Estrategias Evolutivas y RPROP para la Estimación de Redes Neuronales. *Avances en Sistemas e Informática*, Vol. 4, Nro 2, pp. 135–144.
- Palit, A.K.; Popovic, D. [2005] *Computational Intelligence in Time Series Forecasting*. Springer.
- Weigent, A. S.; Rumelhart, D. E.; Huberman, B. A. [1991] Generalization by weight-elimination with application to forecasting. *Advances in Neural Information Processing System*, Morgan Kaufmann, San Mateo, CA. No. 3, pp. 875–882.

