

Efecto de autocorrelación en el modelado de series temporales

H.D. NAVONE, H.A. CECCATTO*

IFIR (UNIVERSIDAD NACIONAL DE ROSARIO - CONICET)
Bv. 27 DE FEBRERO 210 BIS - (2000) ROSARIO - ARGENTINA
e-mail: ceccatto@ifir.edu.ar

and H. WAELBROECK

INSTITUTO DE CIENCIAS NUCLEARES (UNAM)
Circ Ext, C.U., Mexico D.F. 04510

La persistencia es una de las características más comunes de las series temporales correspondientes a fenómenos reales. En este trabajo se investiga el proceso de aprendizaje de dinámicas persistentes usando redes neuronales. Se muestra que para series temporales caóticas la red puede quedar atrapada en un mínimo local de la función error durante largos períodos de aprendizaje, lo que está relacionado con la autocorrelación de la misma. Notablemente, en estos casos la transición hacia la fase en donde la red aprende la dinámica de la serie es bastante abrupta. En cambio, para el caso en que la dinámica persistente es ruidosa el proceso de entrenamiento es suave.

Persistence is one of the most common characteristics of real-world time series. In this work we investigate the process of learning persistent dynamics by neural networks. We show that for chaotic times series the network can get stuck for long training periods in a trivial minimum of the error function related to the long-term autocorrelation in the series. Remarkably, in these cases the transition to the trained phase is quite abrupt. On the contrary, for persistent noisy dynamics the training process is smooth.

PACS: 05.45.+b

Introducción

Las redes neuronales han resultado herramientas importantes en el análisis de series temporales.^{1,2} En los últimos años se han aplicado a un gran número de sistemas, incluyendo secuencias generadas numéricamente (mapas caóticos^{3,4}) y series de datos reales (desde datos meteorológicos⁴ hasta indicadores económicos⁵). Más aún, en muchos casos las redes neuronales permiten una reconstrucción satisfactoria de la dinámica intrínseca de series temporales con un número relativamente pequeño de datos.⁶

En las aplicaciones reales es muy frecuente encontrar series temporales que presentan un alto grado de persistencia. La existencia de esta característica en el registro puede provocar correlaciones y regresiones erróneas debido a que es muy simple para la red aprender el

comportamiento general de los datos antes que la verdadera relación entre las variables.

En este trabajo se discute el problema de la reconstrucción de dinámicas persistentes usando redes neuronales. En particular, se muestran los problemas asociados con el entrenamiento de redes neuronales cuando se les presenta una serie temporal de naturaleza caótica y una serie determinística afectada por ruido, ambas con un alto nivel de persistencia.

Series temporales persistentes

Los efectos que produce la presencia de autocorrelación de largo alcance en el proceso de aprendizaje son investigados considerando dos secuencias de registros generadas computacionalmente. La primera de ellas corresponde a un mapa caótico generado a partir del mapa logístico en donde se incluye la

* Autor a quién debe dirigirse la correspondencia.

presencia de un parámetro libre que controla la autocorrelación. Este mapa, que llamaremos Mapa Logístico Persistente (MLP), está definido por la siguiente relación

$$x_t^* = \alpha x_{t-1}^* + (1 - \alpha) x_t \quad (1)$$

con $x_0^* = 0$. Se denomina x_t^* a las iteraciones del MLP; x_t viene dada por la ecuación cuadrática usual $x_t = r x_{t-1} (1 - x_{t-1})$, donde $r = 3.8$ y $x_0 = 0.23489$.¹ Para $\alpha = 0$ el MLP se reduce al mapa logístico, el cual tiene muy poca autocorrelación (Fig. 1), mientras que para $\alpha = 1$ resulta $x_t^* = x_{t-1}^*$ y por lo tanto el rango de autocorrelación es infinito. Modificando el parámetro α se obtienen series temporales con distintos grados de persistencia; en este trabajo se toma $\alpha = 0.99$.

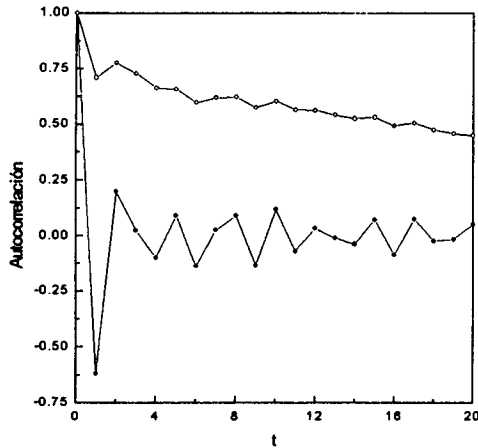


Figura 1: Autocorrelación del mapa logístico (puntos) y del mapa logístico persistente (círculos)

Además del mapa caótico se considera también el proceso de aprendizaje de un registro de naturaleza no-caótica persistente, cuya dinámica está afectada por ruido. Con el propósito de comparar con series reales, se usa un modelo aproximado de la actividad solar medida por el número de Wolf⁷. De acuerdo con este modelo el número promedio de manchas solares para el ciclo n , $f_n(t)$, puede ser aproximado por

$$f_n(t) = \frac{a_n (t - t_{0,n})^3}{\exp\left[\frac{(t - t_{0,n})^2}{b_n^2}\right] - c} \quad (2)$$

donde t se mide en meses. El parámetro a_n es la amplitud del ciclo n , $t_{0,n}$ el tiempo de inicio del ciclo, $c = 0.71$ representa la asimetría de los ciclos, y $b_n \cong 27.12 + 25.15 / (10^3 a_n)^{1/4}$ también medido en meses. Para cada ciclo se utilizan dos parámetros de ajuste: a_n y la duración del ciclo dada por $\Delta_n = t_{0,n+1} - t_{0,n}$. En este trabajo se toman los valores promedios de estos parámetros más un cierto valor de ruido que refleja su distribución estadística sobre 22 ciclos observados. Como la verdadera dinámica de las manchas solares es mucho más compleja que el modelo utilizado, se trabajó con un ruido estadístico 4 veces superior al observado en las distribuciones de los parámetros de ajuste. Esto se hizo justamente para dificultar el aprendizaje de la dinámica del modelo por parte de la red neuronal. La serie temporal resultante se promedió finalmente para obtener un modelo del número promedio de manchas solares anuales con una autocorrelación del orden de la observada.

Aprendizaje de dinámicas persistentes

Para el MLP se consideró una red neuronal 2:2:1 (2 unidades de entrada, 2 unidades ocultas y 1 unidad de salida) la cual es capaz de aprender la dinámica del mapa logístico⁶. Se usó retropropagación de errores como regla de aprendizaje⁸ y varios conjuntos de pesos iniciales generados al azar. Se fijó el parámetro de aprendizaje $\eta = 0.1$ y se tomó igual cantidad de patrones para el conjunto de entrenamiento y validación (entre 50 y 500 patrones). Los datos que conforman los patrones se generaron a partir de (1), descartando las primeras 10^4 iteraciones. El proceso de entrenamiento fue monitoreado siguiendo la evolución de la variancia relativa promedio,

$$ARV(S) = \frac{\sum_{t \in S} (x_t^* - p_t)^2}{\sum_{t \in S} (x_t^* - \bar{x}_t^*)^2} \quad (3)$$

donde S indica alternativamente el conjunto de entrenamiento o validación, p_t es la predicción de la red neuronal para x_t^* , y \bar{x}_t^* es el promedio de x_t^* en S . Esta cantidad es 1 para una predicción constante igual al promedio de las

iteraciones y 0 para una predicción perfecta. En la Fig. 2 se presenta la evolución del ARV en función del número de épocas de entrenamiento (una época corresponde a una presentación del conjunto total de patrones de entrenamiento).

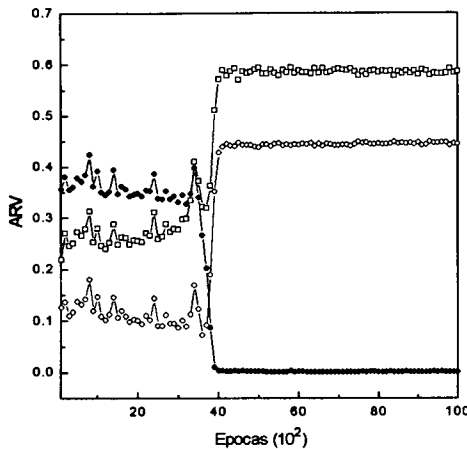


Figura 2: Variancia relativa promedio en función del número de épocas de entrenamiento. Los puntos corresponden al valor de ARV verdadero mientras que los cuadrados y los círculos corresponden a los valores obtenidos desplazando las predicciones uno y dos pasos hacia atrás respectivamente.

Como puede verse, la red neuronal queda atrapada inicialmente en valores altos de ARV , y luego abruptamente aprende la dinámica. A los efectos de entender el comportamiento inicial, se muestran los valores de ARV correspondientes a las predicciones de la red desplazadas uno y dos pasos hacia atrás. El hecho de que inicialmente estos valores sean menores que los correctos indica que la red está prediciendo x_{t+1}^* simplemente como algún promedio de x_t^* y x_{t-1}^* . Además, el ARV de las predicciones de la red desplazadas dos pasos hacia atrás es menor que los valores correspondientes para un desplazamiento de un paso hacia atrás. A partir de la Fig. 1 se concluye que esto es debido a la particularidad de que la autocorrelación de la serie temporal para un desplazamiento temporal de dos pasos es mayor que para un desplazamiento de un paso. También es destacable el hecho de que la transición entre la fase de predicciones triviales a la fase entrenada es remarcablemente abrupta.

En este punto es necesario realizar algunas consideraciones. Primero, no se observó ningún

cambio cualitativo en el proceso de aprendizaje al cambiar el tamaño de los conjuntos de entrenamiento y validación (entre 50 y 500 registros). A nivel cuantitativo se observó una fuerte renormalización del número de épocas requeridas para aprender la verdadera dinámica, creciendo de $O(10^3)$ para 500 patrones a $O(10^4)$ para 50 patrones. Además, como era esperable, se observó que los valores de ARV finales resultaban menores a medida que se incrementaba la cantidad de registros involucrados en el proceso de aprendizaje. Segundo, aproximadamente en el 20% de los experimentos de entrenamiento la red neuronal nunca alcanzó a aprender la dinámica correcta, quedando atrapada en un mínimo local trivial de la función de error. Esto sucede más frecuentemente para los conjuntos de entrenamiento más chicos.

Se realizó un estudio similar para la serie temporal de manchas solares. En este caso se usó una red neuronal 5:3:1 con un parámetro de aprendizaje mucho menor ($\eta=10^{-4}$) para hacer más notorios los efectos de la autocorrelación. Se consideró en primer lugar la serie de manchas solares reales (1700-1991), la cual fue previamente investigada usando redes neuronales.^{2,9} De los 287 patrones disponibles se tomaron los primeros 144 para entrenamiento y el resto para validación. Los resultados de un experimento típico se muestran en la Fig. 3a. Estos resultados deben ser comparados con los correspondientes a la simulación del promedio anual según el modelo (2) con un alto nivel de ruido (Fig. 3b). Ambas curvas se parecen mucho y poseen un comportamiento muy diferente al observado para el MLP. Aún considerando que las series temporales de manchas solares correspondientes al modelo tienen una autocorrelación inicial muy alta (0.88) el proceso de aprendizaje es muy suave. Ahora bien, la región trivial dominada por la autocorrelación está también presente. Diferentes niveles de ruido reducen o aumentan esta región, pero la red nunca queda atrapada en un mínimo local. Es interesante remarcar que este comportamiento sustenta la idea de una dinámica ruidosa no-caótica para las manchas solares.⁶

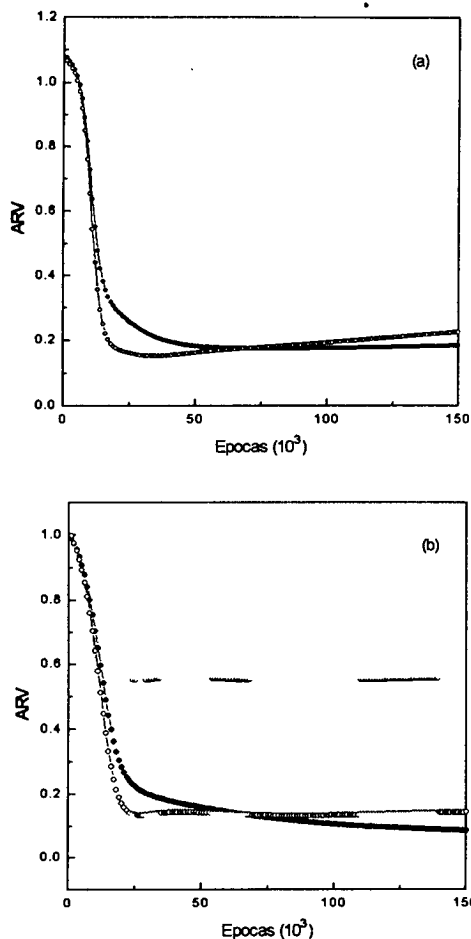


Figura 3: Variancia relativa promedio en función de las épocas de entrenamiento. a) Registro real del promedio anual de manchas solares; b) promedio anual obtenido a partir del modelo (2). Los puntos corresponden al valor correcto de ARV, mientras que los círculos se obtienen al desplazar las predicciones de la red neuronal un paso hacia atrás.

Conclusiones

Se discutieron los efectos de una autocorrelación de largo alcance sobre los procesos de reconstrucción de la dinámica intrínseca de series temporales usando redes neuronales. Se mostró que la red aprende inicialmente a predecir el próximo registro esencialmente con el valor actual, lo que es una consecuencia de la alta autocorrelación de las series. Los ejemplos considerados indican que para series caóticas persistentes la evolución a partir del mínimo trivial de la función error hacia la verdadera dinámica es muy abrupta, mientras que para

sistemas ruidosos no caóticos esta transición es muy suave. En el caso de series temporales caóticas el número de épocas requeridos para aprender la dinámica correcta depende fuertemente de la longitud del registro, con un incremento de aproximadamente dos órdenes de magnitud en los ejemplos considerados. Mas aún, en algunos experimentos la red nunca aprende la dinámica correcta. Esto podría explicar el comportamiento observado frecuentemente en las aplicaciones reales cuando los datos son escasos, en donde la red queda atrapada siempre en el mínimo trivial. En estos casos el hecho de que la red nunca pueda aprender la dinámica real se debe probablemente a la presencia de un atractor de dimensión alta, el cual requiere muchos puntos para ser reconstruido. Para el MLP, dado que el atractor es de baja dimensión, aún con pocos puntos la dinámica puede ser reconstruida. No obstante, los efectos derivados de la persistencia están presentes.

Finalizando, es importante destacar que los diferentes comportamientos observados en el proceso de aprendizaje de dinámicas caóticas y ruidosas podrían ser usados para distinguir caos de ruido en sistemas reales.

Referencias

- 1- Elsner, J.B. & Tsonis, A.A. *Bulletin American Meteorological Society*, **73**, 49-60 (1992).
- 2- Weigend, A.S., Huberman, B.A., & Rumelhart, D.E. *International Journal of Neural Systems*, **1**, 193-209 (1990).
- 3- Elsner, J.B. *Journal of Physics A: Mathematical and General*, **25**, 843-850 (1992).
- 4- Navone, H.D. & Ceccatto, H.A. *Climate Dynamics*, **10**, 305-312 (1994).
- 5- *Neural networks in the capital markets*, Ed. Apostolos-Paul Refenes (Wiley finance editions, 1995).
- 6- Navone, H.D. & Ceccatto, H.A. *Journal of Physics A: Mathematical and General*, **28**, 3381-3388 (1995).
- 7- Hathaway, D.H., Wilson, R.M. & Reichmann, E.J. *Solar Physics*, **151**, 177-190 (1994).
- 8- Rumelhart, D.E., Hinton, G.E. & Williams, R.J. *Nature*, **323**, 533-536 (1986).
- 9- Calvo, R.A., Ceccatto, H.A. & Piacentini, R.D. *The Astrophysical Journal*, **444**, 916-921 (1995).