

INTRODUCCIÓN DE LA INFORMACIÓN DE LAS DERIVADAS DE UNA FUNCIÓN EN EL ENTRENAMIENTO DE REDES NEURONALES

Marcelino Lázaro, Ignacio Santamaría, Carlos Pantaleón

Dpto. Ing. Comunicaciones, ETSII y Telecom, Universidad de Cantabria
Avda. Los Castros, 39005 Santander, Spain
e-mail: {marce,nacho,carlos}@gtas.dicom.unican.es

RESUMEN

En esta comunicación se realiza un estudio sobre la posibilidad de introducir la información de las derivadas de una función en el entrenamiento de las redes neuronales, y las ventajas que se obtienen al hacer uso de esta información. El estudio se ha llevado a cabo sobre varios tipos de redes neuronales, incluyendo las dos más conocidas, el MLP y la RBF, en espacios de entrada unidimensionales y bidimensionales, y se ha comprobado que la utilización de las derivadas en el entrenamiento proporciona ciertas ventajas, especialmente al trabajar en entornos de ruido.

1. INTRODUCCIÓN

El problema general de la reconstrucción de una determinada función, $f(x)$, a partir de una serie de realizaciones de la misma, $f(x_n)$, es un problema muy común y que ha recibido una gran atención en multitud de disciplinas. Sin embargo, existen algunas aplicaciones en las que no es suficiente obtener una aproximación lo más fiel posible de la función original, sino que además es necesario obtener una aproximación razonable de las derivadas de la misma. Entre estas aplicaciones se pueden citar los trabajos de Jordan [1] en robótica, la economía, la dinámica de sistemas caóticos, o el modelado de dispositivos de microondas cuando se quieren tener en cuenta los efectos de intermodulación [2].

Por otro lado, las Redes Neuronales son una alternativa de modelado no lineal que ha despertado un gran interés en los últimos años. Una gran parte de este interés se debe a que las Redes Neuronales son aproximadores universales. Asimismo, es conocido que las Redes Neuronales pueden también aproximar una función y sus derivadas con una precisión arbitraria [3]. Sin embargo, a pesar de esta reconocida capacidad teórica, en la práctica no son comunes las soluciones constructivas que utilizan Redes Neuronales para resolver los problemas de aproximación de una determinada función y sus derivadas. Las soluciones que se emplean en este sentido están, en general, limitadas a las obtenidas a través de la aproximación de la propia función, confiando en la capacidad de la red para que realice una generalización razonable en la estima de las derivadas, pero no se suele utilizar la propia información de las derivadas de la función a la hora de realizar el entrenamiento de las mismas.

En esta comunicación se presentan algunos resultados novedosos obtenidos al introducir la información de las derivadas de la función en el entrenamiento de redes neuronales. Se han utilizado

Este trabajo ha sido financiado parcialmente por la Comunidad Europea y el Gobierno Español a través del proyecto 1FD97-1863-C02-01

las dos Redes Neuronales más conocidas, el *Perceptrón Multicapa* (MLP) y la *Red de Funciones de Base Radial* (RBF) además del *Modelo Lineal a Tramos Suavizado* (MLTS) [4].

2. INFORMACIÓN DE LAS DERIVADAS EN EL ENTRENAMIENTO DE REDES NEURONALES

La introducción de la información de las derivadas en el entrenamiento de las Redes Neuronales se ha llevado a cabo a través de la función de coste a utilizar. En general, el entrenamiento de las Redes Neuronales se realiza a través de la minimización de una determinada función de coste mediante el algoritmo de entrenamiento específico de la red neuronal. La función de coste que se emplea con más asiduidad es el error cuadrático sobre las muestras de la función

$$C = \sum_n \left(f(x_n) - \hat{f}(x_n) \right)^2 \quad (1)$$

donde $f(x_n)$ son las muestras de la función a aproximar y $\hat{f}(x_n)$ son las estimas que proporciona la red neuronal. La alternativa más sencilla para incluir la información de las derivadas en el entrenamiento de la red consiste en introducir en la función de coste un término que pondere el error de la estima de las derivadas. En principio se va a añadir únicamente la información de la derivada primera, con lo que la función de coste es ahora

$$C_d = \lambda_0 \sum_n \left(f(x_n) - \hat{f}(x_n) \right)^2 + \lambda_1 \sum_n \left(f^{(1)}(x_n) - \hat{f}^{(1)}(x_n) \right)^2 \quad (2)$$

donde $f^{(1)}(x_n)$ y $\hat{f}^{(1)}(x_n)$ son, respectivamente, las muestras de la derivada de la función a aproximar y las estimas de la misma. Los parámetros λ_0 y λ_1 se encargan de ponderar la importancia relativa que se va a dar al error sobre las muestras de la señal y de la derivada, y su valor dependerá de aspectos como la aplicación o la fidelidad que se tenga en las muestras de función y derivada. Esta expresión es para el caso simple de espacios de entrada 1D. La extensión al caso de espacios nD , donde se tienen muestras de la derivada primera en distintas direcciones es obvia.

3. RESULTADOS

En este apartado se muestran algunos de los resultados obtenidos con varias Redes Neuronales, el MLP, la RBF y el MLTS, en cu-

yo entrenamiento se ha introducido una función de coste como la descrita en (2). Se ha evaluado la capacidad de aproximación tanto sobre la función como sobre las derivadas, y se ha comparado con el caso en que se entrenaban utilizando solamente la información de la función. Se van a presentar resultados para espacios de entrada $1D$ y $2D$. En el caso $1D$, las funciones de test utilizadas son funciones limitadas en banda compuestas por la combinación lineal de 100 sinusoides de amplitudes, frecuencias y fases aleatorias. En la Figura 1 se muestran los resultados obtenidos con una red MLTS de 10 neuronas, entrenando con las funciones de coste (1) y (2), en función del número de instantes de medida. La relación señal a ruido (SNR), tanto en las muestras de la función como de la derivada es de 10 dB. La figura muestra la relación señal a error (SER) en dB para la reconstrucción de la señal y de la derivada utilizando las funciones de coste dadas por (1) y (2) respectivamente. Los parámetros λ_0 y λ_1 se seleccionaron como $\lambda_0 = 1$ y λ_1 como la relación de potencia entre la función y la derivada.

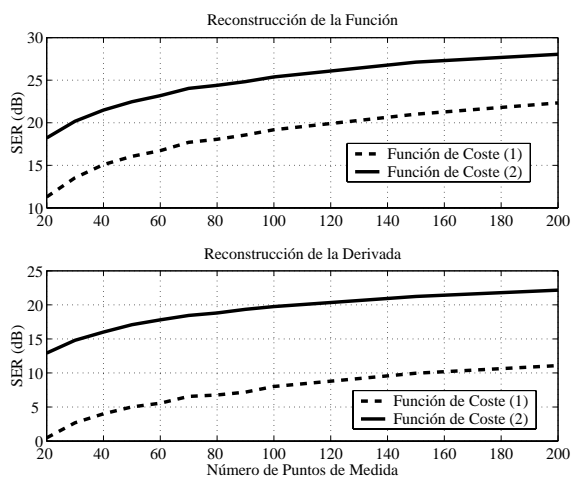


Figura 1. Resultados de reconstrucción en $1D$

Se puede ver que con la función de coste (2) se obtienen unos resultados claramente superiores, especialmente en la reconstrucción de la derivada, donde se ganan más de 10 dB. Es interesante ver que la mejora se produce incluso cuando se tiene el mismo número total de muestras, es decir cuando al utilizar la función de coste (2) se tienen la mitad de instantes de medida que al usar la función (1). Prácticamente los mismos resultados se obtienen con las redes MLP y RBF.

También se han obtenido resultados en espacios de entrada $2D$. En este caso las funciones de test utilizadas son un conjunto de 8 funciones de diferentes características utilizadas en [5] por Cherkassky para la comparación de métodos de modelado adaptativo. Estas 8 funciones suponen un conjunto de test significativo para evaluar resultados. La Figura 2 muestra los resultados obtenidos en la reconstrucción usando la red RBF, utilizando diferentes varianzas en cada dirección del espacio de entrada.

De nuevo se aprecia una importante mejora en los resultados utilizando la función de coste que incluye las derivadas. Usar las derivadas puede llevar a necesitar un número menor de muestras totales (de la función y derivadas con respecto a cada una de las direcciones del espacio de entrada) para obtener los mismos resultados, ya que con 1/3 de puntos de medida se obtienen mejores

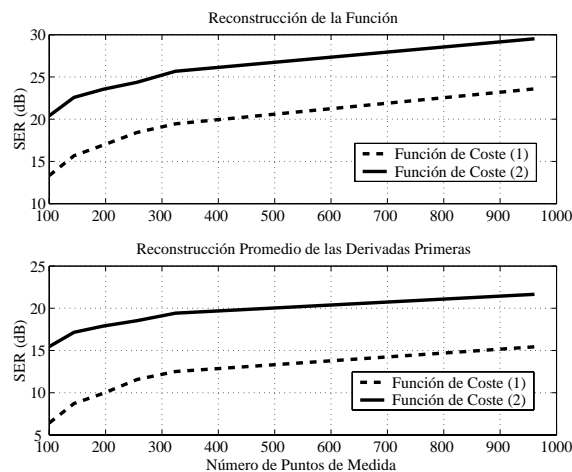


Figura 2. Resultados de reconstrucción en $2D$

resultados que usando sólo información de la función. De nuevo se obtienen resultados muy similares con las otras dos redes, el MLP y el MLTS.

4. CONCLUSIONES

En esta comunicación se ha comprobado que la utilización de las derivadas de la función en el entrenamiento de redes neuronales puede mejorar de forma sensible la aproximación de las mismas. Además, en entornos ruidosos, aun cuando no se esté explícitamente interesado en la información de las derivadas, su utilización puede reducir el número de medidas totales necesarias para obtener los mismos resultados de reconstrucción de la función. La mejora en los resultados es apreciable incluso cuando la relación SNR en las muestras de las derivadas es algo peor que en las de la función.

5. REFERENCIAS

- [1] Jordan, M., "Generic constraints on under specified target trajectories", *Proc. of the IJCNN*, vol 1, pp. 593-606, 1989.
- [2] Crossmun, A. M., Maas, S.A., "Minimization of Intermodulation Distortion in GaAs MESFET small-signal Amplifiers", *IEEE Trans. on MTT*, 7(9):1411-1497, 1989.
- [3] Hornik, K. et al., "Universal Approximation of an Unknown Mapping and its Derivatives using Multilayer Neural Networks", *Neural Networks*, vol 3, pp 551-560, 1990
- [4] Lázaro, M., Santamaría, I., Pantaleón, C., "Smoothing the Canonical Piecewise Linear Model: An Efficient and Derivable Large-Signal Model for MESFET/HEMT Transistors". *IEEE Transactions on Circuits and Systems I*, 48(2):184-192, 2001.
- [5] Cherkassky, V. et al., "Comparison of Adaptive Methods for Function Estimation From Samples", *IEEE Trans. on Neural Networks*, 7(4):969-984, 1996.