

Aprendizaje Multitarea en Problemas con un Número Reducido de Datos

Andrés Bueno-Crespo

Dpto. Informática de Sistemas

Universidad Católica San Antonio de Murcia

e-mail: abueno@pdi.ucam.edu

José-Luis Sancho-Gómez

Dpto. Tecnologías de la Información y Comunicaciones

Universidad Politécnica de Cartagena

e-mail: josel.sancho@upct.es

Abstract- MultiTask Learning (MTL) is a procedure to train a neural network to learn several related tasks simultaneously considering one of them as main task and the others as secondary tasks. In this paper, we have tested a method to obtain artificially tasks which are related with the main one, because in many real cases, knowledge about problem to be solved is uncertain. We use sample selection techniques to generate related tasks with the main one, in particular, samples close the classification boundary. Moreover, a new procedure to train MultiLayer Perceptrons with generated tasks is described.

I. INTRODUCCIÓN

El Aprendizaje Multitarea (MultiTask Learning - MTL) permite entrenar una tarea junto con un conjunto de tareas relacionadas con ésta. Esto permite que estas tareas secundarias ayuden a la principal a mejorar su entrenamiento.

Aquí proponemos utilizar procedimientos de selección de muestras que nos permitirán obtener tareas secundarias de forma artificial para los casos en que queramos utilizar un esquema de Aprendizaje Multitarea y no se dispone de tareas relacionadas con la principal.

II. APRENDIZAJE MULTITAREA

El Aprendizaje Multitarea permite entrenar Redes Neuronales Artificiales (RNA) aprendiendo varias tareas simultáneamente, considerando una de ellas como tarea principal y las otras como secundarias. La información que proporcionan estas tareas secundarias es utilizado por la principal como sesgo.

En este tipo de esquemas han trabajado R. Caruana [1-4] y Y. Bengio [5] proponiendo alternativas en los algoritmos y estudios sobre la importancia de compartir dominios al utilizar este esquema de Aprendizaje Multitarea.

En la figura 1 podemos ver las diferencias entre un esquema de entrenamiento simple y uno multitarea. Como se puede observar, en el esquema multitarea se comparten las capas ocultas con lo que esa información proporcionada por las tareas secundarias van a ayudar a la principal en su aprendizaje.

Desafortunadamente, no siempre disponemos de tareas relacionadas o estas tareas no son muy claras de identificar

para poder utilizar un esquema MTL. En estos casos, proponemos utilizar procedimientos de selección de muestras para obtener estas tareas secundarias relacionadas con la principal.

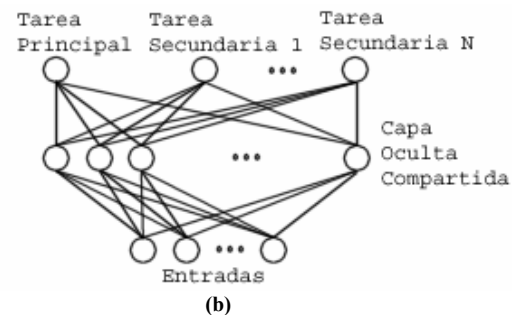
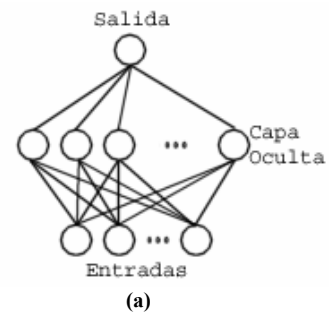
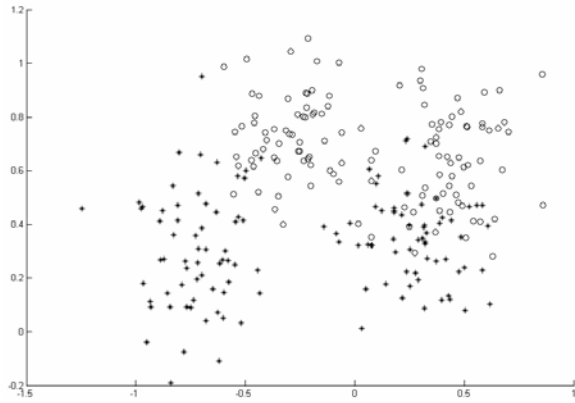


Fig. 1. (a) Esquema de Aprendizaje Simple (STL). (b) Esquema de Aprendizaje Multitarea (MTL). Todas las tareas comparten la capa oculta. Las tareas secundarias ayudan a la principal en su aprendizaje.

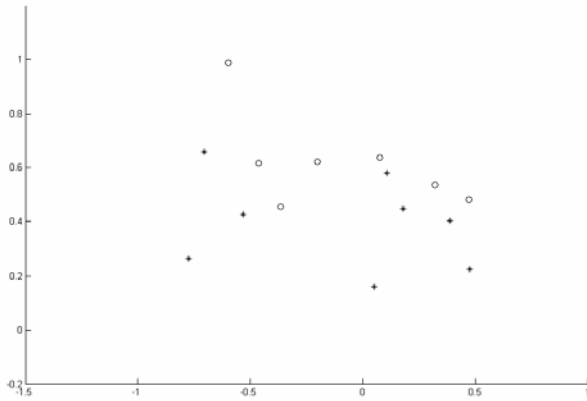
III. CREACIÓN ARTIFICIAL DE TAREAS

Nuestro objetivo es crear tareas artificiales relacionadas con la principal. Para ello utilizaremos un procedimiento de selección de muestras llamado Edición de Datos (Data Editing). Este procedimiento fue desarrollado para reducir el conjunto de entrenamiento y las grandes cantidades de almacenamiento, ya que en algunos problemas sólo necesitamos una pequeña fracción del conjunto de entrenamiento para obtener una buena aproximación del contorno de decisión del clasificador K-NN. Para nuestro propósito, utilizaremos la combinación del algoritmo {K,L}-NN y el algoritmo de Condensación-NN. El algoritmo {K,L}-NN se utiliza para eliminar la zona de datos

superpuestos, mientras que el algoritmo de Condensación-NN lo utilizamos para obtener un conjunto de datos destacado del conjunto de muestras que nos servirá para construir las tareas secundarias.



(a)



(b)

Fig. 2. (a) Conjunto de muestras. (b) Conjunto de muestras seleccionadas usando el procedimiento de Edición de Datos.

Como se puede apreciar, el conjunto editado (Fig. 2.b) es una pequeña fracción del conjunto original formado por las muestras que forman el contorno habiendo eliminado las superpuestas. Ahora, la RNA tiene dos tareas a llevar a cabo, la principal consiste en aprender el conjunto completo de entrenamiento y la tarea secundaria consiste en encontrar la frontera del mismo conjunto de entrenamiento pero dando más importancia al conjunto de muestras seleccionado.

Supongamos que tenemos un problema de clasificación de dos clases (la extensión a más clases es obvia). Una vez que la tarea secundaria es obtenida artificialmente, la red se entrena utilizando el algoritmo de retropropagación, pero con algunas modificaciones orientadas a implementar un esquema MTL.

Modificamos el error cuadrático para una muestra \mathbf{x}^n multiplicándolo por una constante β_k tal que:

$$E^n = \frac{1}{2} \sum_{k=1}^c \beta_k (y_k^n - t_k^n)^2 \quad (1)$$

donde $c-1$ es el número de tareas secundarias relacionadas con la principal, y_k es la salida de la muestra k , y t_k es el correspondiente valor deseado para el patrón de entrada \mathbf{x}^n .

β_k presenta dos posibles valores, dependiendo de si k es la tarea principal (T_p) o la secundaria (T_s):

$$\beta_k = \begin{cases} 1, & \mathbf{x}^k \in T_p \\ \alpha, & \mathbf{x}^k \in T_s \end{cases} \quad (2)$$

Así, el error para la muestra n y dos salidas (principal y secundaria) viene dado por:

$$E^n = \frac{1}{2} [(y_1 - t_1)^2 + \alpha (y_2 - t_2)^2] \quad (3)$$

Para enfatizar el conjunto de datos frente al conjunto de muestras de la tarea secundaria, proponemos modificar el valor de α :

$$\alpha = \begin{cases} 1, & \mathbf{x}^n \in \text{Conjunto Editado} \\ \lambda, & \mathbf{x}^n \notin \text{Conjunto Editado} \end{cases} \quad (4)$$

donde $\lambda > 1$.

con estas modificaciones, los gradientes en el algoritmo de retropropagación son:

$$\frac{\partial E^n}{\partial w_{ji}} = \delta_j x_i \quad (5)$$

$$\frac{\partial E^n}{\partial w_{kj}} = \delta_k z_j \quad (6)$$

donde

$$\delta_j = \alpha z_j (1 - z_j) \sum_{k=1}^c w_{kj} (y_k - t_k) \quad (7)$$

siendo z_j una función no lineal de activación de la muestra j .

IV. EXPERIMENTO

Para testear nuestro algoritmo, utilizaremos un conjunto de datos utilizado por B. D. Ripley [8]. Se trata de un problema de clasificación de dos dimensiones con una probabilidad de error teórica de 0.08. La figura 2.a representa este conjunto.

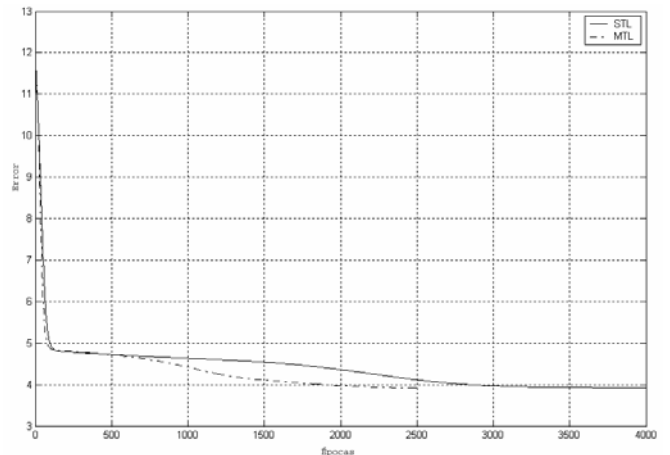


Fig. 3. Entrenamiento MTL vs. STL. Las tareas secundarias ayudan a la principal a mejorar su aprendizaje. MTL mejora el número de épocas en un 37,5 %.

El conjunto de entrenamiento y el de validación están formados por 250 muestras mientras que para el conjunto de test se han utilizado 1000 muestras, dos entradas, 1 salida y 3 neuronas en la capa oculta.

La media de la probabilidad de error y la desviación estándar para 10 repeticiones en un sistema STL es: 0.090 ± 0.003 . Para mejorar este resultado usaremos un sistema MTL. Suponemos que no tenemos tareas relacionadas, por lo que usaremos Edición de Datos para obtener un grupo reducido de muestras significativas que representará la tarea secundaria.

Para cada valor de λ , se ha repetido el entrenamiento 10 veces, con inicializaciones aleatorias de los pesos. Los mejores resultados se han obtenido con valores de λ entre 0.6 y 0.9 (Tabla 1).

Esquema	λ	Prob. Error \pm Desv. std
STL	-	0.090 ± 0.003
MTL	0.9	0.082 ± 0.002
	0.8	0.082 ± 0.001
	0.7	0.081 ± 0.002
	0.6	0.082 ± 0.003
	0.5	0.083 ± 0.004
	0.4	0.084 ± 0.003
	0.3	0.085 ± 0.003
	0.2	0.086 ± 0.005
	0.1	0.087 ± 0.003
	0.0	0.088 ± 0.002

Tabla 1. Media de la probabilidad de error y desviación estándar para diferentes valores de λ en un esquema MTL frente a un esquema STL.

V. CONCLUSIONES

Si se dispone de un conjunto de tareas relacionadas entre sí, se puede utilizar un esquema de Aprendizaje Multitarea (MTL) para entrenar una red neuronal considerando una de ellas como tarea principal y las otras como secundarias. En la mayoría de los casos, no se dispone de estas tareas relacionadas. En este artículo hemos testeado un método para obtener artificialmente tareas relacionadas con la principal. Para ello, proponemos seleccionar muestras del conjunto original de datos, en particular, las muestras que forman el borde de separación de las clases, habiendo quitado previamente las muestras superpuestas. La Edición de Datos es un ejemplo de procedimiento de selección que produce un conjunto reducido de muestras de interés. En este trabajo, el proceso de Edición de Datos ha sido implementado por los algoritmos $\{K,L\}$ -NN y Condensación-NN. Se ha entrenado una Red Neuronal Artificial utilizando un esquema de Aprendizaje Simple y uno de Aprendizaje Multitarea creando tareas secundarias con el procedimiento propuesto de Edición de Datos. De acuerdo con los resultados obtenidos, creemos que utilizar la Edición de Datos para obtener tareas relacionadas con la principal es un método recomendado para mejorar el aprendizaje cuando no tenemos tareas relacionadas para utilizar un esquema MTL.

RECONOCIMIENTO

Este trabajo está subvencionado por el Ministerio de Educación y Ciencia, otorgado por TIC2002-03033.

REFERENCIAS

- [1] R. Caruana, S. Baluja and T. Mitchell, "Using the Future to 'Sort Out' the Present", Rankprop and Multitask Learning for Medical Risk Prediction. Advances in Neural Information Processing System 8, (Proceeding of NIPS-95), 1996, pp. 959-965.
- [2] R. Caruana, "Algorithms and Applications for Multitask Learning". International Conference on Machine Learning, 1996, pp. 87-95
- [3] R. Caruana, "Multitask Learning". Machine Learning, 28, 1997, pp. 41-75
- [4] R. Caruana, "Learning Many Related Tasks at the Same Time with Backpropagation". Advanced in Neural Information Processing System 7, (Proceeding of NIPS-94), 1995, pp. 656-664.
- [5] J. Ghosh and Y. Bengio, "Bias Learning, Knowledge Sharing". IEEE Transactions on Neural Networks, Vol. 14, N° 4, July 2003.
- [6] D. J. Hand and B. G. Batchelor, "An Edited Condensed Nearest Neighbour Rule". Information Sciences 14, 1978, pp. 171-180
- [7] P. E. Hart, "The Condensed Nearest Neighbour Rule". IEEE Transactions on Information Theory 14, 1968, pp. 515-516
- [8] B. D. Ripley, "Pattern Recognition and Neural Networks". Cambridge University Press, 1996.