

Comparação entre duas técnicas para o treinamento de Perceptron de Múltiplas Camadas

Fábio Dall Cortivo¹, Ezzat S. Chalhoub², Haroldo F. de Campos Velho²

¹Programa de Doutorado em Computação Aplicada – CAP
Instituto Nacional de Pesquisas Espaciais – INPE

²Laboratório Associado de Computação e Matemática Aplicada – LAC
Instituto Nacional de Pesquisas Espaciais – INPE

{fabio.cortivo, ezzat, haroldo}@lac.inpe.br

Abstract. *This work presents a comparison between two techniques of training a Multi-Layer Perceptron artificial neural network. We solve a problem in Hydrologic Optics, in which we estimate the single scattering albedo, where a neural network acts as the inverse operator. One of the training techniques used is the backpropagation algorithm error and the other is the quasi-Newton method. The network training with the quasi-Newton method present better results when compared with the network training using backpropagation algorithm error.*

Resumo. *Este trabalho faz uma comparação entre duas técnicas para o treinamento de uma rede neural artificial do tipo Perceptron de Múltiplas Camadas. Para isso, faz-se uso de um problema em Ótica Hidrológica, no qual se pretende estimar o albedo de espalhamento simples, em que a rede neural atua como o operador inverso. Uma das técnicas de treinamento utilizada é o algoritmo de retropropagação do erro e a outra é o método quasi-Newton. A rede treinada com o método quasi-Newton apresentou resultados superiores em comparação com a rede treinada com o algoritmo de retropropagação do erro.*

Palavras-chave: *Redes Neurais, Transferência Radiativa (ETR), Ótica Hidrológica, Back-propagation, método quasi-Newton*

1. Introdução

O algoritmo de retropropagação do erro é, talvez, a técnica mais popular para o treinamento de Perceptron de Múltiplas Camadas (MLP), e baseia-se no gradiente da função erro. Porém, este algoritmo pode apresentar uma convergência lenta, ou então, ficar preso em mínimos locais¹ [Haykin 2001, Fausett 1994]. Diante desses problemas, e de certas dificuldades encontradas em classificar corretamente certos padrões, em um trabalho já realizado [Cortivo et al. 2010], optou-se por usar outra técnica de treinamento, no entanto, mantendo o mesmo problema para que a comparação entre as técnicas seja válida.

A regra de atualização para as variáveis da rede, que será utilizada neste trabalho, é baseada no método quasi-Newton [Dennis and Moré 1977]. Este método usa mais informações acerca do problema para efetuar a atualização nas variáveis da rede e, além disso, possibilita a inclusão de cada parâmetro das funções de ativação nas variáveis a serem otimizadas.

¹Sabe-se que atingir um ponto de mínimo em otimização não linear, não é garantia de que se tenha atingido o ponto de mínimo global.

O problema utilizado é um problema em Ótica Hidrológica, no qual se pretende estimar o albedo de espalhamento simples. Neste caso, a rede utilizada atuará como o operador inverso.

2. Problema Abordado

A Equação de Transferência Radiativa (ETR) é um modelo matemático que descreve a interação do campo de radiação em um ambiente que absorve, emite e espalha essa radiação. A ETR para problemas estacionários, em geometria plano-paralela, em uma determinada faixa de frequência, com dependência azimutal e sem a presença de termo fonte, tem a forma

$$\mu \frac{\partial}{\partial \tau} I(\tau, \vec{\xi}) + I(\tau, \vec{\xi}) = \frac{\varpi_0}{4\pi} \int_{-1}^1 \int_0^{2\pi} \beta(\vec{\xi}; \vec{\xi}') I(\tau, \vec{\xi}') d\vec{\xi}', \quad (1)$$

em que τ representa a espessura ótica, $\vec{\xi} = \vec{\xi}(\mu, \phi)$ representa um vetor unitário com variação polar $\mu = \cos \theta$, ($0 \leq \theta \leq \pi$) e azimutal $0 \leq \phi \leq 2\pi$, $\beta(\vec{\xi}; \vec{\xi}')$ é chamada de função de fase e representa a probabilidade de um feixe de radiação incidindo, com intensidade $I(\tau, \vec{\xi})$, na direção $\vec{\xi}$, sofrer um espalhamento na direção $\vec{\xi}' = (\mu', \phi')$, com $d\vec{\xi}' = d\phi' d\mu'$ é o elemento diferencial, e ϖ_0 é o albedo de espalhamento simples.

Para completar a formulação do problema é necessário que sejam definidas às condições de contorno, as quais representam o fluxo na entrada e na saída do ambiente. Estas condições são dadas por

$$I(0, \vec{\xi}) = G_+(\vec{\xi}) \quad \text{para } \mu > 0, \quad \text{e} \quad I(\tau_0, \vec{\xi}) = G_-(\vec{\xi}) \quad \text{para } \mu < 0. \quad (2)$$

Detalhes de como obter a solução do problema acima definido podem ser encontrados em [Cortivo et al. 2010, Dall Cortivo 2008, Chalhoub 2005].

3. Redes Neurais

As RNAs são sistemas computacionais estruturados em uma aproximação à computação baseada em ligações, e a inspiração desses sistemas proveio da tentativa de modelar o funcionamento do cérebro humano em um computador. Os precursores no desenvolvimento do neurônio artificial foram o neurofisiologista Warren McCulloch e o matemático Walter Pitts [McCulloch and Pitts 1943]. O neurônio artificial consiste em um combinador linear e uma função de ativação, com isso, tenta modelar o funcionamento de um neurônio do cérebro humano. Para isso, o vetor de valores \vec{x} são as entradas das sinapses conectadas a um neurônio, através dos w_j , que são os pesos, representando a “energia” das conexões, b é um parâmetro de ajuste e φ é a função de ativação, que tem por finalidade restringir a amplitude da saída do neurônio. A Figura 1(a) mostra uma representação gráfica do neurônio artificial. A representação em forma de equação, para a saída de um neurônio, é dada por

$$y = \varphi(v) = \varphi \left(\sum_{j=1}^n w_j x_j + b \right), \quad (3)$$

em que φ é a função de ativação, x_j são os valores de entrada do neurônio, b é um viés associado ao neurônio, e w_j são os pesos sinápticos do neurônio artificial.

Um MLP é construído conectando neurônios uns aos outros de modo que formem uma rede, como mostra a Figura 1(b). Esta estrutura é constituída de uma camada de entrada, uma ou mais camadas ocultas de nós de fonte e uma camada de saída, em que a função de ativação de cada neurônio/camada oculto(a) pode ser linear ou não. Além disso, a estrutura do MLP é capaz de resolver problemas não lineares, diferentemente do perceptron que só é capaz de resolver problemas linearmente separáveis [Haykin 2001].

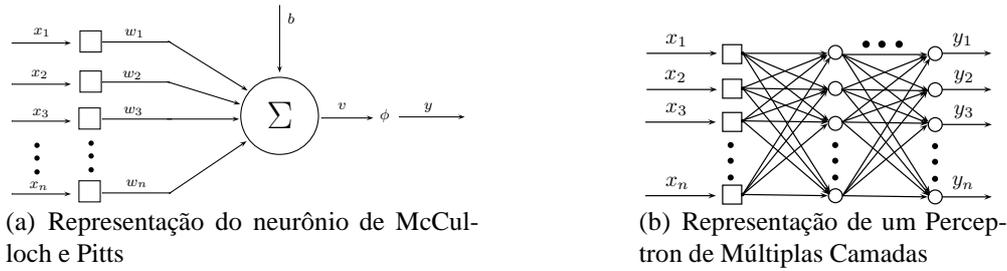


Figura 1. Representação do neurônio de McCulloch e Pitts e de um MLP

Como funções de ativação não lineares que podem ser usadas em função do problema e/ou a critério do pesquisador são: as funções sigmoide e a tangente hiperbólica representadas, respectivamente, por

$$\varphi(v) = \frac{1}{1 + e^{-av}} \quad \text{e} \quad \varphi(v) = \frac{1 - e^{-av}}{1 + e^{-av}} \quad \text{em que } a > 0. \quad (4)$$

Além dessas duas há a função linear $\varphi(v) = av$, em que $a \in \mathbb{R}$, entre outras que podem ser utilizadas [Haykin 2001].

4. Metodologia

Descreve-se nestas duas primeiras seções a arquitetura de cada rede utilizada em cada problema. Posteriormente são feitos breves comentários acerca da geração dos padrões de treinamento e de validação, e por fim, como se deu o treinamento da rede utilizando o método proposto neste trabalho.

4.1. Metodologia Anterior

A rede utilizada no problema já resolvido continha 20 unidades de entrada, as quais representaram o tamanho do vetor de entrada, uma camada oculta com 40 neurônios e uma camada de saída com apenas um neurônio. A função de ativação utilizada na camada de saída foi a função linear, $\varphi(v) = av$ com $a = 1$, e na camada oculta, foi utilizada a função sigmoide com parâmetro $a = 3$ para os primeiros trinta neurônios, e $a = 6$ para os dez últimos neurônios. Além disso, utilizou-se o algoritmo de retropropagação do erro com uma atualização dos pesos sinápticos baseada na regra delta generalizada, definida por [Haykin 2001]

$$w_{ji}^{(l)}(n+1) = w_{ji}^{(l)}(n) + \alpha \left[w_{ji}^{(l)}(n-1) \right] + \eta \delta_j^{(l)}(n) y_i^{(l-1)}(n), \quad (5)$$

onde l representa a camada, $w_{ji}^{(l)}$ é o peso sináptico do neurônio j da camada l que é alimentado pelo neurônio i da camada $l - 1$, $y_i^{(l-1)}(n)$ é o sinal de saída do neurônio

i na camada anterior $l - 1$ na iteração n , η é chamado de taxa de aprendizagem e α é chamada de constante de momento. Outros detalhes relacionados com o problema já resolvido podem ser encontrados em [Cortivo et al. 2010], e com a regra de aprendizagem em [Haykin 2001].

4.2. Metodologia Atual

Pelo motivo da regra adotada para a atualização das variáveis, método quasi-Newton, e pelo fato de se ter incluído o parâmetro de cada função de ativação como variáveis a serem otimizadas, a estrutura da rede mudou significativamente. A rede utilizada nesta segunda metodologia consiste em dez unidades de entrada, uma camada oculta com apenas onze neurônios e uma camada de saída com apenas um neurônio. Já as funções de ativação, continuaram sendo a função sigmoide e a função linear.

O treinamento da rede é supervisionado, porém, a apresentação dos padrões é feita no modo lote, ou seja, somente é feita a atualização das variáveis após todos os padrões de treinamento terem sido apresentados a rede, diferentemente do algoritmo de retropropagação do erro, no qual, o algoritmo também é do tipo supervisionado, no entanto, a correção das variáveis é feita a cada padrão de treinamento apresentado.

Observa-se que no problema já resolvido, trinta neurônios possuíam o mesmo valor para o parâmetro a da função de ativação, ou seja, possuíam exatamente a mesma função de ativação. O mesmo é válido para os outros dez neurônios restantes, os quais possuíam outro valor de a , porém, o mesmo para todos. Em resumo, pode-se dizer que a camada oculta possuía duas funções de ativação. Já neste problema, cada função de ativação tem seu próprio valor de a , ou seja, cada neurônio terá a sua função de ativação.

Referências

- Chalhoub, E. S. (2005). Discrete-ordinates solution for uncoupled multi-wavelength radiative transfer problems. *Journal of Quantitative Spectroscopy & Radiative Transfer*, 92:335 – 349.
- Cortivo, F. D., Chalhoub, E. S., Silva, J. D. S., and Campos Velho, H. F. (2010). Estimativa do albedo de espalhamento simples usando uma rede neural de múltiplas camadas. In *Anais...*, pages 411 – 417, São Carlos, São Paulo. XXXIII Congresso Nacional de Matemática Aplicada e Computacional, Sociedade Brasileira de Matemática Aplicada e Computacional.
- Dall Cortivo, F. (2008). Reconstrução de termo fonte e condições de contorno em óptica hidrológica com séries de fourier.
- Dennis, J. E. and Moré, J. J. (1977). Quasi-newton methods, motivation and theory. *Society for Industrial and Applied Mathematics Review*, 19(1):46 – 89.
- Fausett, L. (1994). *Fundamentals of neural networks*. Prentice Hall International, New Jersey.
- Haykin, S. (2001). *Redes neurais*. Bookman, Porto Alegre.
- McCulloch, W. and Pitts, W. H. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5:115–133.