

CLASSIFICAÇÃO DE EEG COM REDES NEURAIS ARTIFICIAIS UTILIZANDO ALGORITMOS DE TREINAMENTO DO TIPO *EXTREME LEARNING MACHINE* E *BACK-PROPAGATION*

Tavares S. Tatiana*, Francisco A. O. Nascimento * e Cristiano J. Miosso **

* Departamento de Engenharia Elétrica, Universidade de Brasília, Brasília, Brasil

** Programa de Pós-Graduação em Engenharia Biomédica (UnB/FGA), Gama, Brasil

e-mail: tavares@aluno.unb.br

Resumo: Este artigo apresenta a classificação de sinais eletroencefalográficos (EEG) utilizando redes neurais artificiais (RNA). Serão avaliados os resultados obtidos nos testes de classificação dos sinais de dois grupos de pacientes, um grupo saudável e outro grupo com epilepsia. A avaliação tem o intuito de auxiliar em diagnóstico médico para a detecção de epilepsia e analisar o desempenho dos classificadores neurais. Para tal propósito foram implementados dois algoritmos, máquina de aprendizado extremo (ELM, *extreme learning machine*) e retropropagação (BP, *back-propagation*) para o treinamento da RNA. Os sinais de entrada são apresentados à rede com e sem a extração prévia de características. O algoritmo ELM mostrou um melhor desempenho em relação ao BP, alcançando 94,90% de acurácia no teste em 0,0079 segundos de treinamento.

Palavras-chave: redes neurais artificiais, máquina de aprendizado extremo, rede neural sem realimentação, retropropagação.

Abstract: This article presents the classification of electroencephalographic signals (EEG) using artificial neural networks (ANN). They will be evaluated the results in grading tests the signs of two groups of patients, a healthy group and another group with epilepsy. The evaluation is intended to assist in medical diagnosis for epilepsy detection and analyze the performance of neural classifiers. For this purpose were implemented two algorithms, extreme learning machine (ELM, *extreme learning machine*) and back propagation (BP, *backpropagation*) for training of RNA. The input signals are presented to the network with or without prior extraction characteristics. The ELM algorithm showed better performance in relation to BP, reaching 94.90% accuracy in test 0.0079 seconds training.

Keywords: artificial neural networks, extreme learning machine, feedforward neural networks, back-propagation.

Introdução

O eletroencefalograma (EEG) é uma importante ferramenta para o diagnóstico e tratamento de pacientes com doenças neurológicas, tais como crises epiléticas [1] [2], depressão [3], doença de Parkinson [4] [5] e deficiências motoras [6] [7] [8].

O EEG tem sido cada vez mais utilizado em diferentes contextos, no estudo [9] é realizada uma análise comparativa entre os sistemas de classificação análise discriminante linear (LDA, *Linear Discriminant analysis*), máquina de vetor de suporte (SVM, *Support Vector Machine*) e máquina de aprendizado extremo (ELM, *Extreme Learning Machine*) aplicados a interface cérebro máquina (BCI, *brain computer interface*). O ELM é proposto por [10] para estimar a vigilância de operadores de máquina e seu desempenho é comparado com SVM e ELM modificado. Em [11] os autores propõem duas técnicas de classificação, ELM e SVM para o reconhecimento de emoções a partir de sinais de EEG. O ELM é também abordado por [12] para a detecção de epilepsia, onde a extração prévia de características é realizada pela transformada wavelet. No trabalho [13] os classificadores ELM, SVM e retropropagação (BP, *Back-propagation*) são comparados na realização da tarefa de detecção de epilepsia.

Neste trabalho é proposta uma rede neural artificial RNA [14], sem realimentação com uma camada oculta e sobre tal arquitetura é aplicado os algoritmos de treinamento ELM e BP, com o propósito de classificação de sinais de EEG de um grupo de pacientes saudáveis e outro grupo com epilepsia. A acurácia e o tempo de resposta de cada classificador foram analisados. Nesta abordagem destacamos o fato de que os dados de entrada são apresentados à RNA com e sem a extração prévia de características. Três configurações, para os dados de entrada, serão apresentadas à RNA, uma configuração com extração prévia de características, havendo uma redução na dimensionalidade dos dados de entrada e duas configurações sem a extração prévia de característica, nas quais o tamanho das amostras apresentadas à entrada da rede é o mesmo, diferenciando-se apenas quanto à informação espectral.

Materiais e métodos

Todo o desenvolvimento dos algoritmos foi realizado em linguagem e plataforma *Matlab*. Para uso de BP foram utilizadas as funções existentes na *toolbox* de redes neurais e para o ELM fez-se necessária a implementação do algoritmo conforme proposto por [15].

Banco de dados – No início deste trabalho já se dispunha de um banco de dados da Universidade de Bonn [16]. Cada arquivo é composto de 100 séries temporais com duração de 23,6 segundos e frequência amostral de 173,61 Hz. A base de dados original é dividida em cinco conjuntos: Z, O, N, F, S. Os conjuntos Z e O correspondem a séries de EEG extra cranial, de cinco voluntários saudáveis (com olhos abertos e fechados, respectivamente). As outras três séries são de pacientes epiléticos, com a série S obtida durante um período de crise, e as séries F e N (com eletrodos localizados na zona epileptogênica e hipocampo oposto, respectivamente) durante períodos livres de crise.

Neste estudo, foram utilizados dois conjuntos; O e S, correspondendo às classes saudável e doente, respectivamente.

Todas as séries devem pertencer a uma classe saudável ou doente. Para este problema 75% e 25% das amostras são escolhidas aleatoriamente para o treino e para o teste em cada ensaio, respectivamente.

A seguir são apresentadas as configurações do banco de dados para a realização da classificação;

Configuração 1: Características - Neste estudo foi feita a extração de um conjunto de 72 características relevantes, energia e valor *RMS*, de cada série temporal em janelas de 4 segundos, antes da aplicação do algoritmo classificador. Há uma redução na dimensionalidade dos dados de entrada, sendo necessário 72 neurônios de entrada da rede.

Configuração 2: Sinal Completo - São as 100 séries temporais com duração de 23,6 segundos e frequência amostral de 173,61 Hz, sendo 4097 neurônios de entrada da rede.

Configuração 3: Sinal Completo FFT - Para esta simulação foi calculado o módulo da transformada rápida de Fourier (FFT, *Fast Fourier Transform*), detalhada em [17], do sinal de entrada completo, sendo 4097 neurônios de entrada da rede. Nesta configuração é apresentada à rede as informações de amplitudes do sinal.

Nas configurações 2 e 3 o sinal de entrada tem o mesmo tamanho, e o número de exemplos apresentados às redes é o mesmo para todas as configurações.

Feedforward Neural Network – Uma *Feedforward Neural Network* é uma RNA na qual a informação se move em uma única direção, para frente, a partir dos nós (neurônios) de entrada, através de nós ocultos (se houver) e para os nós de saída. Não há ciclos ou *loops* na rede, o que a torna mais simples sendo um dos primeiros tipos de RNA's criadas. Não possuindo, pois, realimentação da saída para a entrada, tais redes são consideradas sem memória, estes são, provavelmente, os tipos mais populares de redes [18].

Uma rede neural sem realimentação com uma camada oculta (SLFN, *Single-hidden Layer Feedforward Network*) se trata de uma RNA sem *loops*, contendo apenas uma camada oculta, é a arquitetura

sobre a qual se pretende aplicar o algoritmo de ELM e o BP para o propósito deste trabalho.

BP – O processo de treinamento de uma rede neural consiste basicamente em apresentar pares entrada-saída à rede, calcular as saídas da rede em função das entradas, calcular o erro entre a saída desejada e a saída calculada e alterar os valores sinápticos por algum tipo de algoritmo. À medida que os pares entrada-saída vão sendo apresentados e os pesos das sinapses atualizados, o erro médio quadrático (MSE, *Mean Square Error*) da saída da rede tende a diminuir [19]. Existem alguns tipos mais importantes de algoritmos, dentre os quais o mais utilizado e eficiente é o chamado de BP. O processo de treinamento da rede neural pelo método de BP realiza, como o próprio nome diz, a retro propagação dos erros calculados das saídas em direção às entradas. O erro calculado na saída de um neurônio é multiplicado pela derivada da função de ativação daquele neurônio e propagado para a sua entrada. Este valor então é enviado para todos os neurônios da camada anterior pesado pelas respectivas sinapses [20]. Neste trabalho o algoritmo de BP utiliza o método do gradiente descendente para corrigir os valores sinápticos.

A taxa de aprendizagem é uma constante que controla a velocidade de convergência do algoritmo, ou seja, a rapidez com que os valores sinápticos levam o MSE para um mínimo local ou global na superfície de erro do problema. Para pequenos valores de taxa de aprendizagem a atualização das sinapses é mais lenta, fazendo com que se percorra com maior precisão a superfície de erro do problema. Para valores altos de taxa de aprendizagem, apesar da convergência ocorrer de maneira mais rápida, pode acontecer oscilações no treinamento, podendo às vezes até impossibilitar a convergência para um mínimo local ou global. O BP é um algoritmo fortemente sensível à variação do número de neurônios da camada oculta.

Para a implementação do BP foi realizada a validação cruzada, que busca obter a melhor generalização da rede. Um conjunto de 75% das amostras são escolhidas aleatoriamente para encontrar os melhores valores da taxa de aprendizagem e do número de neurônios da camada oculta, na validação cruzada. Foram testados diversos valores de épocas para o treinamento da rede tendo como objetivo a melhor eficiência nos resultados.

ELM – foram originalmente desenvolvidas para SLFN's. Sua essência é de que, diferentemente do entendimento de aprendizagem, a camada oculta de uma SLFN não precisa ser ajustada. A implementação típica de uma ELM aplica nós de computação aleatória, que devem ser independentes dos dados de treinamento. De acordo com a teoria tanto o erro de treinamento quanto a norma dos pesos precisam ser minimizados. Uma vez que para as ELM's os pesos das camadas ocultas não precisam ser ajustados e seus parâmetros podem ser fixos, os pesos das saídas podem ser resolvidos pelo método dos quadrados mínimos [15].

A SLFN com K nós ocultos, com função de ativação $g(x)$ pode aproximar N amostras (x_i, t_i) com erro zero. Uma SLFN pode ser representada matricialmente por

$$H\beta = T \quad (1)$$

$$H_{N \times K} = \begin{bmatrix} g(w_1 \cdot x_1 + b_1) & \cdots & g(w_K \cdot x_1 + b_K) \\ \vdots & \cdots & \vdots \\ g(w_1 \cdot x_N + b_1) & \cdots & g(w_K \cdot x_N + b_K) \end{bmatrix} \quad (2)$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_K^T \end{bmatrix}_{K \times m} \quad T = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix}_{N \times m} \quad (3)$$

onde w é um vetor de pesos que conecta os nós ocultos e os nós de entrada, β é o vetor de pesos que conecta os nós ocultos com os nós de saída e b (*bias*) é o limiar dos nós ocultos [15] [21].

H é a matriz de saída da camada oculta. Se $K = N$, então a matriz H é quadrada e inversível quando o vetor de pesos w_i e os *bias* são escolhidos aleatoriamente. Nesta condição, a SLFN pode aproximar as amostras de treinamento com o erro zero.

Na prática, $K \ll N$, fazendo com que H não seja uma matriz quadrada e possa ser que não exista w , *bias* e β que satisfaça $H\beta = T$. Dessa forma, a solução para esse problema é calcular:

$$\hat{\beta} = H^+ T \quad (4)$$

onde H^+ é a matriz inversa generalizada de Moore-Penrose [15].

Treinamento da SLFN utilizando ELM:

Dado: um conjunto de treinamento $N = \{(x_i, t_i) \mid x_i \in \mathbb{R}^n, t_i \in \mathbb{R}^m, i = 1, \dots, N\}$, a função de ativação $g(x)$, infinitamente diferenciável, e o número de nós na camada oculta K .

Passo 1: Gerar aleatoriamente os pesos w e os *bias*

Passo 2: Calcular a matriz de saída da camada oculta (H);

Passo 3: Calcular os pesos da camada de saída $\hat{\beta} = H^+ T$
Onde $T = [t_1, \dots, t_N]^T$

Métricas – Para a realização experimental dos algoritmos, utilizou-se como base as seguintes métricas: acurácia no treinamento e no teste e o tempo de treinamento e de teste.

Resultados

Mil ensaios foram realizados e os resultados da classificação dos sinais de EEG para um grupo de pacientes saudáveis e epiléticos são apresentados nas Tabelas 1, 2 e 3. A função de ativação para os dois algoritmos é sigmoide e o ELM é implementado com 20 neurônios na camada oculta. O número de neurônios da

camada oculta do BP e a taxa de aprendizagem são selecionados na validação cruzada com o objetivo de alcançar o menor MSE.

Tabela 1: Performance da configuração 1 – Características. BP com 25 neurônios na camada oculta e taxa de aprendizagem 0,01.

Algoritmo	Treino		Teste	
	Time (s)	Acurácia (%)	Time (s)	Acurácia (%)
ELM	0.0022	94,73	2,5e-4	90,62
BP	13,5	99,32	0,047	100

Tabela 2: Performance da configuração 2 – Sinal Completo. BP com 25 neurônios na camada oculta e taxa de aprendizagem 0,02.

Algoritmo	Treino		Teste	
	Time (s)	Acurácia (%)	Time (s)	Acurácia (%)
ELM	0.0079	77,40	0,022	74,01
BP	5,516	53,02	0.078	54

Tabela 3: Performance da configuração 3 – Sinal Completo com FFT. BP com 21 neurônios na camada oculta e taxa de aprendizagem 0,015.

Algoritmo	Treino		Teste	
	Time (s)	Acurácia (%)	Time (s)	Acurácia (%)
ELM	0.0079	97,04	0,002	94,90
BP	48,656	99,33	0,047	98

Discussão

Na Tabela 1 observamos que com a extração prévia de características o BP alcançou 100% de acurácia no teste em 13,5 segundos de treinamento. O ELM se destaca, nesta configuração quanto a sua rapidez na classificação, realizando o treinamento em 0.0022 segundos para uma acurácia no teste de 90,62%.

Os resultados apresentados na Tabela 2 foram gerados com o intuito de avaliar os algoritmos sem a extração prévia de características. O ELM alcançou uma maior acurácia, mas inferior aos resultados da Tabela 1.

A fim de melhorar os resultados sem a extração prévia de característica, foram levantados os resultados da Tabela 3. Os dois algoritmos apresentam acurácia no teste acima de 90%, sendo o ELM mais rápido no treinamento.

Considerando as três configurações propostas para os dados de entrada, o algoritmo BP teve acurácia no teste que alcançou 100%, mas com o tempo de treinamento longo em comparação ao ELM. O ELM se mostrou capaz de realizar a classificação do EEG com uma acurácia de 94,90% no teste, em um tempo de

treinamento de 0,0079 segundos, sendo assim uma importante ferramenta no auxílio a diagnóstico médico.

Agradecimentos

Agradecimentos à CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) pelo fomento financeiro a esta pesquisa.

Referências

- [1] Verma, N. et al. A micro-power eeg acquisition soc with integrated feature extraction processor for a chronic seizure detection system. *IEEE Journal of Solid-State Circuits*, v. 45, n. 4, p. 804816, April 2010. ISSN 0018-9200.
- [2] Chisci, L. et al. Real-time epileptic seizure prediction using ar models and support vector machines.
- [3] Niemiec, A. J.; Lithgow, B. J. Alpha-band characteristics in eeg spectrum indicate reliability of frontal brain asymmetry measures in diagnosis of depression. In: 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference. [S.l.: s.n.], 2005. p. 75177520. ISSN 1094-687X.
- [4] Handojoseno, A. M. A. et al. An eeg study of turning freeze in parkinson's disease patients: The alteration of brain dynamic on the motor and visual cortex. In: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). [S.l.: s.n.], 2015. p. 661876621. ISSN 1094-687X.
- [5] Handojoseno, A. M. A. et al. The detection of freezing of gait in parkinson's disease patients using eeg signals based on wavelet decomposition. In: 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. [S.l.: s.n.], 2012. p. 6972. ISSN 1094-687X.
- [6] Shedeed, H. A.; Issa, M. F.; El-Sayed, S. M. Brain eeg signal processing for controlling a robotic arm. In: Computer Engineering Systems (ICCES), 2013 8th International Conference on. [S.l.: s.n.], 2013. p. 152157.
- [7] Victorino, J. et al. Improving eeg-bci analysis for low certainty subjects by using dictionary learning. In: Signal Processing, Images and Computer Vision (STSIVA), 2015 20th Symposium on. [S.l.: s.n.], 2015. p. 17.
- [8] Pfurtscheller, G. et al. Separability of eeg signals recorded during right and left motor imagery using adaptive autoregressive parameters. *IEEE Transactions on Rehabilitation Engineering*, v. 6, n. 3, p. 316325, Sep 1998. ISSN 1063-6528.
- [9] Sarah N. Carvalho, Thiago B.S. Costa, Luisa F.S. Uribe, Diogo C. Soriano, Glauco F.G. Yared, Luis C. Coradine, Romis Attux, Comparative analysis of strategies for feature extraction and classification in SSVEP BCIs, *Biomedical Signal Processing and Control*, Volume 21, August 2015, Pages 34-42, ISSN 1746-8094.
- [10] Li-Chen Shi, Bao-Liang Lu, EEG-based vigilance estimation using extreme learning machines, *Neurocomputing*, Volume 102, 15 February 2013, Pages 135-143, ISSN 0925-2312.
- [11] R. E. J. Yohanes, W. Ser and G. b. Huang, Discrete Wavelet Transform coefficients for emotion recognition from EEG signals, *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, San Diego, CA, 2012, pp. 2251-2254.
- [12] Yuan Q, Zhou W, Zhang J, Li S, Cai D, Zeng Y. EEG classification approach based on the extreme learning machine and wavelet transform. *Clin EEG Neurosci*. 2012 Apr;43(2):127-32.
- [13] Yuan Q, Zhou W, Li S, Cai D. Epileptic EEG classification based on extreme learning machine and nonlinear features. *Epilepsy Res*. 2011 Sep;96(1-2):29-38.
- [14] Haykin, S. Redes neurais principios e prática. [S.l.]: Porto Alegre: Bookman, 2001. (900p).
- [15] Huang, G.-B.; Zhu, Q.-Y.; Siew, C.-K. Extreme learning machine: Theory and applications. *Neurocomputing*, v. 70, n. 1â3, p. 489 501, 2006. ISSN 0925-2312. Neural NetworksSelected Papers from the 7th Brazilian Symposium on Neural Networks (SBRN '04)7th Brazilian Symposium on Neural Networks.
- [16] Andrzejak, R. et al., 2001. Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. *Physical Review E*, 64(6), p.061907.
- [17] Oppenheim, A. V.; Schaffer, R. W., *Discrete-Time Signal Processing*, Prentice Hall, [1989, 1999, 2009]
- [18] Srinivasan, V.; Eswaran, C.; Sriiram, N. Approximate entropy-based epileptic eeg detection using artificial neural networks. *IEEE Transactions on Information Technology in Biomedicine*, v. 11, n. 3, p. 2887295, May 2007. ISSN 1089-7771.
- [19] Srinivasan, B.; Prasad, U. R.; Rao, N. J. Back propagation through adjoints for the identification of nonlinear dynamic systems using recurrent neural models. *IEEE Transactions on Neural Networks*, v. 5, n. 2, p. 2137228, Mar 1994. ISSN 1045-9227.
- [20] Purnamasari, P. D.; Ratna, A. A. P.; Kusumoputro, B. Eeg based patient emotion monitoring using relative wavelet energy feature and back propagation neural network. In: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). [S.l.: s.n.], 2015. p. 28202823. ISSN 1094-687X.
- [21] Zhu Q.; Siew, C. Extreme learning machine: a new learning scheme of feedforward neural networks. 2004.