

Avaliação de Desempenho da Rede Neural Máquina de Aprendizado Extremo na Detecção de Crises Epilépticas

J. Peixoto*, N. Rodrigues*, G. A. Barreto*, R. Becco**

*Universidade Federal do Ceará (UFC), Depto. Engenharia Teleinformática (DETI), Fortaleza, Ceará

**Universidade Federal do Ceará (UFC), Departamento de Clínica Médica, Fortaleza, Ceará

e-mails: {juliojrps, natanrs21}@gmail.com, gbarreto@ufc.br, rodrigo.becco@ebserh.gov.br

Abstract - *In this paper we evaluate the performance of the Extreme Learning Machine (ELM) network in the task of EEG-based epileptic seizures detection. Our goal is to investigate how well this machine learning algorithm, which belongs to the class of randomized neural networks, performs for two standard feature extraction methods. A comprehensive performance evaluation is carried out, with the results corroborating our hypothesis that the performance of the ELM is strongly dependent on the feature extraction method used.*

Palavras-chave: *crises epilépticas, densidade espectral de potência, coeficientes LPC, máquina de aprendizado extremo, classificação.*

Introdução

Crises epilépticas correspondem a estados cerebrais em que se observam descargas anormais, excessivas e síncronas, de neurônios localizados majoritariamente no córtex cerebral. Esta atividade anormal, que é intermitente e geralmente dura de alguns segundos a poucos minutos, afeta cerca de 1% da população mundial [1].

O sinal de eletroencefalograma (EEG) é utilizado clinicamente de forma rotineira para diagnosticar, monitorar e localizar focos epileptogênicos. Tal sinal consiste em uma representação do somatório das atividades síncronas de um conjunto de milhões de neurônios que tem uma orientação espacial semelhantes. Além disso, sinais de EEG são altamente ruidosos, de natureza não estacionária e de dinâmica altamente não linear [2]. Tais características tornam o tratamento de sinais de EEG (e.g. para fins de detecção de crises convulsivas) como uma tarefa muito desafiadora, apesar de todo o recente desenvolvimento nas áreas de processamento não linear de sinais e de aprendizado de máquina. A literatura sobre a aplicação de classificadores não-lineares na detecção/classificação de convulsões, como MLP e SVM, é extensa [3].

Mais recentemente tem havido interesse crescente na utilização de um modo alternativo de projeto de redes neurais artificiais multicamadas, que *grosso modo* envolve a randomização dos pe-

sos da primeira camada da rede. Destaque é dado à máquina de aprendizado extremo (ELM, sigla em Inglês) [4], cujo o desempenho em tarefas de reconhecimento de padrões que envolvam o processamento de sinais de EEG vem sendo avaliada em um número cada vez maior de trabalhos (ver, por exemplo, as referências [5–7]).

Do exposto, neste artigo abordamos uma lacuna na literatura sobre ELM para processamento de sinais de EEG. Lacuna esta que consiste em estudar o efeito que a escolha dos métodos de extração de atributos do sinal de EEG tem sobre a acurácia do classificador baseado na rede ELM. Para este fim, selecionamos a tarefa de detecção de crises epilépticas [8]. Os métodos de extração de atributos escolhidos são os seguintes: (i) método de Welch para estimação da densidade espectral de potência (PSD, sigla em Inglês) e (ii) coeficientes de codificação linear preditiva (LPC, sigla em Inglês). O segundo método é tipicamente usado em tarefas de processamento de voz/fala, mas resolvemos utilizá-lo aqui no processamento de sinais de EEG. Os resultados obtidos são promissores.

Materiais e Métodos

O problema de interesse, detecção de crises epilépticas geralmente é tratado como um problema de classificação binária. Neste sentido, o classificador deve diferenciar entre períodos de atividade epiléptica normal ou convulsiva.

O banco de dados utilizado neste estudo foi produzido pelo *Children's Hospital de Boston*, Massachusetts, EUA, em parceria com o MIT, e está disponível para uso público¹. Este banco contém sinais de EEG de 24 pacientes já diagnosticados com epilepsia, onde o monitoramento de vídeo EEG foi realizado após a retirada da medicação anticonvulsivante [9]. Os sinais têm taxa de amostragem de 256 Hz, o conjunto de eletrodos do escalpo (não invasivo) segue o padrão 10-20 e o número de canais varia de 23 a 32. Ao todo,

¹CHB-MIT Scalp EEG Database <http://www.physionet.org/pn6/chbmit/>

197 convulsões foram catalogadas em 141 arquivos, com 196 minutos de intervalos contendo crises epilépticas para todos os pacientes.

Uma vez que a EEG é uma série temporal complexa, com comportamento aperiódico, não estacionário e não linear, para uso em classificação de padrões faz-se necessário extrair atributos que preservem informações relevantes contidas nos sinais originais. Conforme já mencionado, escolhemos duas técnicas para este fim, a saber, o periodograma de Welch [10] e coeficientes LPC [11]. É importante ressaltar, porém, que ambos os métodos assumem estacionariedade dos sinais a serem processados. Neste caso, para lidar com a não-estacionariedade do sinal de EEG, a sequência original é segmentada em subsequências menores, que são então assumidas estacionárias.

Periodograma de Welch - Este método é usado para estimar a PSD de um sinal estocástico por meio da média dos periodogramas das subsequências menores. Como resultado, uma estimativa mais suavizada da PSD, sem as flutuações aleatórias resultantes do próprio processo da estimação, é obtida.

Mais especificamente, o sinal de EEG original de cada canal é dividido em um número K de possíveis segmentos sobrepostos. Uma função de ponderação v_n , também chamada de janela, é definida em cada um desses segmentos e os periodogramas correspondentes são calculados e depois é calculada a média. Se $x_n^{(k)}$ representa a amostra x_n da k -ésima subsequência de dados (de comprimento L), então o periodograma modificado para essa subsequência é calculado como

$$\hat{P}_x^{(k)} = \frac{1}{L} \left| \sum_{n=0}^{L-1} v_n x_n^{(k)} e^{-j\omega n} \right|^2, \quad k = 1, \dots, K \quad (1)$$

onde $\omega = 2\pi f$ (em rad/s) é a frequência angular, e a janela v_n deve obedecer a seguinte propriedade: $(1/L) \sum_{n=0}^{L-1} v_n^2 = 1$. Então, a estimação da PSD do sinal, para cada frequência ω , é dada por

$$\hat{S}_x(\omega) = \frac{1}{K} \sum_{k=1}^K \hat{P}_x^{(k)}(\omega) \quad (2)$$

Coefficientes LPC - Este método pressupõe que a dinâmica da k -ésima subsequência de dados pode ser descrita por um modelo autorregressivo de ordem p , $AR(p)$:

$$x_n^{(k)} = a_{k,1}x_{n-1}^{(k)} + a_{k,2}x_{n-2}^{(k)} + \dots + a_{k,p}x_{n-p}^{(k)} + \epsilon_n, \quad (3)$$

em que $a_{k,1}, \dots, a_{k,p}$ são os coeficientes do modelo $AR(p)$ para a k -ésima subsequência, e ϵ_n segue uma distribuição do tipo ruído branco aditivo gaussiano (AWGN), isto é, $\epsilon_n \sim N(0, \sigma_\epsilon^2)$. Os coeficientes $\{a_{k,l}\}_{l=1}^p$ associados a cada subsequência $k = 1, \dots, K$, podem ser estimados por vários métodos, tais como o método de Yule-Walker ou mínimos quadrados ordinários (MQO) [11].

Montagem dos Vetores de Atributos - Consideremos um arquivo de um certo paciente com $N = 23$ canais. Cada canal provê um sinal com aproximadamente 1 hora de duração, amostrado a uma taxa de 256Hz. Cada canal é processado em intervalos consecutivos de 2s que não se sobrepõem (denominados de épocas e simbolizados pelo letra T). Para a taxa de amostragem dada, cada época T contém 512 amostras. Para um sinal de 1h de duração, temos $3600/2 = 1800$ épocas ao todo.

▷ **Método 1 (Welch)**: Adaptamos a metodologia proposta em [9] de modo a usar como atributos a média da PSD ao longo de todos os canais para cinco faixas de ritmos cerebrais². Assim, para os $N = 23$ canais de um dado paciente, executamos os seguintes passos.

- **Passo 1** - Dentro da época T atual, aplicar o método de Welch ao segmento correspondente em cada um dos N canais.
- **Passo 2** - Converter os valores de PSD resultantes para decibéis (dB).
- **Passo 3** - Segmentar a PSD de cada canal em $M = 5$ bandas de frequência que cobrem o intervalo de 0,5 a 128 Hz, correspondentes às faixas de frequências dos ritmos cerebrais.
- **Passo 4** - Para um dado canal, calcular a média da PSD em cada faixa de frequência e, em seguida, a média daquela faixa de frequência ao longo de todos os canais.
- **Passo 5** - Montar vetor de atributos de dimensão $M = 5$, referente à época atual, da seguinte forma: $\mathbf{x}_T = [\delta_T, \theta_T, \alpha_T, \beta_T, \gamma_T]$.
- **Passo 6** - Fazer $T = T + 1$, e repetir os passos anteriores.

Ao final, teremos 1800 vetores de atributos devidamente rotulados para treinar e testar o classificador de interesse.

² δ (0,5-4Hz), θ (4-7,5Hz), α (8-13Hz), β (14-30Hz) e γ (≥ 30 Hz)

▷ **Método 2 (LPC)**: A construção do vetor de atributos via coeficientes do LPC também evoluiu em épocas. No entanto, em vez de especificar o número M de bandas de frequência, precisamos especificar a ordem do modelo $\text{AR}(p)$. Após alguma experimentação, definimos $p = 4$. Valores maiores não melhoraram consideravelmente a acurácia do classificador, enquanto valores mais baixos levaram a uma degradação em seu desempenho. Assim, para os $N = 23$ canais de um dado paciente, executamos os seguintes passos.

- **Passo 1** - Dentro da época T atual, aplicar a equação de Yule-Walker para estimar os coeficientes do modelo $\text{AR}(p)$. Repetir este procedimento para todos os N canais.
- **Passo 2** - Concatenar os coeficientes estimados para cada um dos N canais do EEG. Este processo forma um vetor de atributos \mathbf{x}_T de dimensão $p \times N = 92$, definido como:

$$\mathbf{x}_T = [a_{1,1} \ \cdots \ a_{1,p} \ | \ \cdots \ | \ a_{N,1} \ \cdots \ a_{N,p}]^\top \quad (4)$$

Para finalizar, ressaltamos que, devido à própria natureza da tarefa, há muito mais vetores de atributos rotulados como *estado normal*, ou seja, em estado de não crise (rótulo -1) do que como *estado epiléptico* (rótulo +1). Na verdade, em torno de apenas 2% dos sinais EEG analisados correspondem a intervalos que contêm convulsões. Isso implica que a tarefa de detecção de crises epilépticas envolve classes altamente desbalanceadas.

Extreme Learning Machine - A rede ELM [4] é uma arquitetura neural *feedforward* do tipo randomizada com uma única camada oculta. Os pesos da entrada para a camada oculta são escolhidos aleatoriamente, enquanto os pesos da camada oculta de saída são determinados analiticamente. A rapidez no projeto de redes randomizadas é uma característica vantajosa, em comparação ao lento processo de treinamento de redes neurais mais tradicionais (e.g. MLP e RBF).

Do ponto de vista da arquitetura, a rede ELM pode ser entendida como uma versão simplificada da RVFL (*Random Vector Functional Link Network*) [12] na qual o caminho linear direto é removido. A saída de rede ELM é calculada como

$$y_n = \mathbf{w}^T \mathbf{h}_n, \quad (5)$$

onde \mathbf{w} é o vetor de pesos do único neurônio de saída³. O vetor de ativação dos neurônios ocultos

na iteração n , $\mathbf{h}_n = \phi(\mathbf{M}\mathbf{x}_n)$, é dado por

$$\mathbf{h}_n = [\phi(\mathbf{m}_1^T \mathbf{x}_n + b_1), \dots, \phi(\mathbf{m}_q^T \mathbf{x}_n + b_q)]^T \quad (6)$$

em que q é o número de neurônios da camada oculta, \mathbf{m}_j é o vetor de pesos do j -ésimo neurônio oculto ($k = 1, \dots, q$), $\phi(\cdot)$ é a função de ativação sigmoideal, e b_j é o limiar do j -ésimo neurônio oculto.

Na estimação de \mathbf{w} optamos por utilizar o método MQO. Desta forma, \mathbf{w} é estimado por

$$\mathbf{w} = (\mathbf{H}\mathbf{H}^T)^{-1}\mathbf{H}\mathbf{d}, \quad (7)$$

onde $\mathbf{H} = [\mathbf{h}_1 \ | \ \mathbf{h}_2 \ | \ \cdots \ | \ \mathbf{h}_{N_1}]$ é uma matriz $q \times N_1$ cujas colunas são as ativações dos q neurônios ocultos em resposta aos N_1 vetores de entrada de treinamento.

Resultados

Nesta seção, relatamos os resultados da análise de desempenho que realizamos. Para lidar com as categorias desequilibradas, equalizamos deliberadamente a proporção de casos positivos a negativos por paciente. Em seguida, dividimos aleatoriamente as instâncias disponíveis por paciente em 3 subgrupos: treinamento (70%), validação (20%) e teste (10%).

Para os testes a serem reportados adiante, utilizamos 80 neurônios na camada oculta. Este hiperparâmetro foi escolhido com base no melhor resultado para o pior cenário dentre todos os pacientes. Para cada paciente, foram realizadas 100 rodadas independentes para os 2 cenários. No final de cada fase de teste, as seguintes figuras de mérito foram computadas a partir da matriz de confusão: acurácia, sensibilidade, especificidade e coeficiente de correlação de Matthews (CCM).

Os resultados numéricos são apresentados na Tabela 1 são de seis indivíduos diferentes (P1 a P6), cujos resultados numéricos em conjuntos de teste são típicos entre os coletados em todo o conjunto de indivíduos em dois cenários distintos. Na tabela, os métodos de extração de atributos são simbolizados pelas letras L (LPC) e W (Welch).

Discussão e Conclusões

Ao analisar a Tabela 1 é possível perceber claramente que o desempenho da rede ELM é melhor ao usar os coeficientes LPC em vez do periodograma de Welch. Este padrão é observado para

³Nosso problema é de classificação binária.

Tabela 1 – Desempenho da rede ELM para os diferentes métodos de extração de atributos.

	Acurácia		Sensibilidade		CCM	
	L	W	L	W	L	W
P1	98,54	96,38	96,50	93,43	0,967	0,918
P2	98,30	98,03	97,80	97,89	0,966	0,961
P3	97,82	97,19	92,71	92,81	0,944	0,927
P4	89,59	84,31	76,35	57,15	0,728	0,572
P5	98,73	95,88	94,04	79,31	0,956	0,851
P6	99,51	98,97	97,04	94,20	0,982	0,962

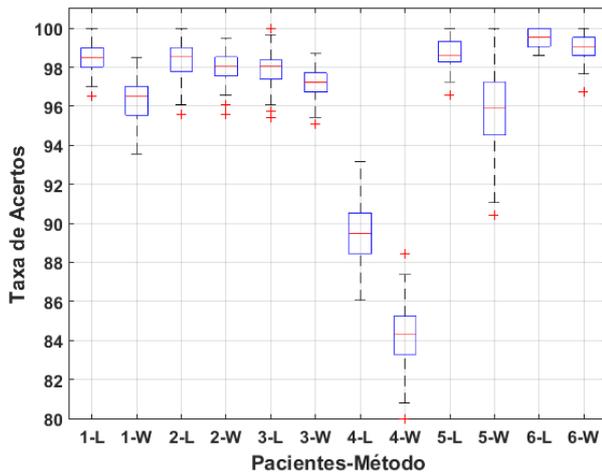


Figura 1 – Boxplots da acurácia da rede ELM para os 6 pacientes e os dois métodos de extração de atributos avaliados (LPC e Welch).

todos os 6 pacientes. Uma outra forma de visualizar os resultados é através dos boxplots para uma dada figura de mérito. Por exemplo, na Figura 1 estão mostrados os resultados para a acurácia de classificação correta para as diferentes combinações paciente-método de extração. Para um dado indivíduo, as medianas da acurácia (linha vermelha) sempre são mais altas quando se usa o método dos coeficientes LPC. Em suma, apesar da facilidade de implementação e projeto, o desempenho da rede ELM depende fortemente do algoritmo de extração de atributos escolhido.

Futuramente, pretendemos estender este trabalho através da utilização de outros métodos de extração de atributos, tais como transformada Wavelet e uso de dinâmica não-linear. Comparação do desempenho da rede ELM com os de outros classificadores não-lineares, tais como MLP e SVM, também está sendo executada.

Agradecimentos: Os autores agradecem ao CNPq (309451/2015-9), CAPES e FUNCAP pelo apoio financeiro.

Referências

- [1] K. Lehnertz, F. Mormann, T. Kreuz, R. Andrzejak, C. Rieke, P. David, and C. Elger, “Seizure prediction by nonlinear EEG analysis,” *IEEE Engineering in Medicine and Biology Magazine*, vol. 22, no. 1, pp. 57–63, 2003.
- [2] D. P. Subha, P. K. Joseph, R. Acharya, and C. M. Lim, “EEG signal analysis: A survey,” *Journal of Medical Systems*, vol. 34, no. 2, pp. 195–212, 2010.
- [3] T. N. Alotaiby, S. A. Alshebeili, T. Alshawi, I. Ahmad, and F. E. A. El-Samie, “EEG seizure detection and prediction algorithms: a survey,” *EURASIP Journal on Advances in Signal Processing*, vol. 2014, no. 1, p. 183, 2014.
- [4] G. Huang, G.-B. Huang, S. Song, and K. You, “Trends in extreme learning machines: A review,” *Neural Networks*, vol. 61, no. 1, pp. 32–48, 2015.
- [5] Y. Wang, Z. Li, L. Feng, C. Zheng, and W. Zhang, “Automatic detection of epilepsy and seizure using multiclass sparse extreme learning machine classification,” *Computational and Mathematical Methods in Medicine*, vol. 2017, no. ID 6849360, pp. 1–10, 2017.
- [6] S. Ding, N. Zhang, X. Xu, L. Guo, and J. Zhang, “Deep extreme learning machine and its application in EEG classification,” *Mathematical Problems in Engineering*, vol. 2015, no. ID 129021, pp. 1–11, 2015.
- [7] H. Zhao, X. Guo, M. Wang, T. Li, C. Pang, and D. Georgakopoulos, “Analyze EEG signals with extreme learning machine based on PMIS feature selection,” *International Journal of Machine Learning and Cybernetics*, pp. 1–7, 2015.
- [8] H. Adeli and S. Ghosh-Dastidar, *Automated EEG-Based diagnosis of neurological disorders: Inventing the Future of Neurology*. New York: CRC Press, 2010.
- [9] A. Shoeb and J. Gutttag, “Application of machine learning to epileptic seizure detection,” in *Proceedings of the 27th International Conference on Machine Learning (ICML’2010)*, pp. 1–8, 2010.
- [10] P. D. Welch, “The use of the fast fourier transform for the estimation of power spectra,” *IEEE Transactions on Audio Electroacoustics*, vol. 15, no. 2, pp. 70–73, 1967.
- [11] C. W. Therrien, *Discrete Random Signals and Statistical Signal Processing*. New Jersey: Prentice-Hall, 1992.
- [12] L. Zhang and P. N. Suganthan, “A comprehensive evaluation of random vector functional link networks,” *Information Sciences*, vol. 367–368, pp. 1094–1105, 2016.