

ESTUDO DE MÁQUINA DE APRENDIZAGEM EXTREMA APLICADA À APROXIMAÇÃO DE FUNÇÃO

RÔMULLO RANDELL MACEDO CARVALHO^{1*}; FRANCISCO ITALLO BRANDÃO RODRIGUES²;
GUILHERME PEREIRA GONÇALVES³; BRENO HENRIQUE SOUSA ANDRADE⁴;
FRANCISCO LEÃO DE OLIVEIRA⁵

¹Graduando em Engenharia Elétrica, bolsista do PET, UFPI, Teresina-PI, randellromullo@gmail.com

²Graduando em Engenharia Elétrica, bolsista do PET, UFPI, Teresina-PI, itallobrandao7@hotmail.com

³Graduando em Engenharia Elétrica, UFPI, Teresina-PI, guilz1@hotmail.com

⁴Graduando em Engenharia Elétrica, UFPI, Teresina-PI, brenohs77@hotmail.com

⁵Graduando em Engenharia Elétrica, UFPI, Teresina-PI, fco.leao@hotmail.com

Apresentado no

Congresso Técnico Científico da Engenharia e da Agronomia – CONTECC'2017

8 a 11 de agosto de 2017 – Belém-PA, Brasil

RESUMO: Redes neurais artificiais são capazes de modelar a função de transferência de processos sem que esta seja realmente conhecida; a computação exigida por estas requer, por vezes, menos processamento que diversos outros processos. A máquina de aprendizagem extrema, derivada da aprendizagem supervisionada, apresenta algumas vantagens sobre redes baseadas puramente em supervisão, como o *perceptron* de múltiplas camadas; a otimização do tempo na modelagem daquela é bastante superior à deste. Esse estudo objetiva implementar o algoritmo *forward* da máquina de aprendizagem extrema para aplicar na aproximação de função, bem como avaliar a importância do principal parâmetro para esta rede: a quantidade de neurônios na camada escondida. Os resultados são apresentados como uma média para 1000 repetições de treino-teste da rede, a fim de apresentar resultados mais confiáveis estatisticamente, uma vez que, em determinados pontos, a rede neural em questão depende bastante da aleatoriedade.

PALAVRAS-CHAVE: Máquina de aprendizado extremo, redes neurais artificiais, aproximação de função, algoritmo *forward*.

EXTREME LEARNING MACHINE STUDY APPLIED TO FUNCTION APPROXIMATION

ABSTRACT: Artificial neural networks are able to model the process transfer function without it being really known; the computation required by these needs, at times, less processing than several other processes. The extreme learning machine, derived from supervised learning, has some advantages over networks purely based on supervision, such as the multilayer perceptron; the optimization of the time in the modeling of it is much superior to this one. This study aims to implement the forward algorithm of the extreme learning machine to apply in the function approach, as well as to evaluate the importance of the main parameter for this network: the amount of neurons in the hidden layer. The results are presented as a mean for 1000 network training-test repetitions in order to present statistically more reliable results, since at certain points, the neural network in question depends a lot on the randomness.

KEYWORDS: Extreme Learning Machine (ELM), artificial neural network, function approximation.

INTRODUÇÃO

Uma rede neural artificial (RNA) é uma máquina virtual (implementada por algoritmo) ou física (por meio de componentes eletrônicos) capaz de modelar, à semelhança do cérebro, determinada tarefa, em geral, de classificação de padrões ou dedução comportamental. Ela é composta por unidades simples denominadas neurônios que funcionam como elementos primordiais de seu processamento e,

por vezes, consegue obter uma resposta mais otimizada que processos convencionais para se obter a dinâmica de sistemas diversos.

Segundo Haykin (2001), McCulloch e Pitts, respectivamente um neuroanatomista e um matemático, foram pioneiros ao publicar seu artigo, em 1943, a respeito da modelagem clássica do neurônio, reunindo seus estudos em neurofisiologia do primeiro e em lógica do segundo. Esse trabalho foi responsável pelo pontapé inicial na construção do conhecimento aplicado de redes neurais artificiais existente hoje, bem como, por si só, demonstrou que uma quantidade definida de neurônios ajustadas de forma tal que as conexões sinápticas operassem em paralelo seria capaz de processar qualquer função computável.

Os trabalhos de McCulloch e Pitts foram minuciosamente estudados, aplicados e expandidos ao longo da segunda metade do século XX, culminando no desenvolvimento de diversas redes neurais com vários neurônios (a principal delas, o perceptron de múltiplas camadas, ou *multilayer perceptron*, *MLP*). A tendência inicial seguida deu origem à aprendizagem supervisionada, em que se utiliza de estímulo de entrada e resposta esperada correspondente para treinar à rede até que esta seja capaz de identificar padrões por si só.

Antes mesmo de surgir a principal aprendizagem de máquina em oposição à supervisionada, os mapas auto-organizáveis de Kohonen (ou aprendizagem não supervisionada), em 1982, deu-se início a um estudo interessante na década de 1960, derivado da supervisão no treinamento de redes neurais: as máquinas de aprendizagem, ou *learning machines*. Estas se baseiam na separabilidade linear de padrões por meio de hipersuperfícies.

A máquina de aprendizagem extrema, ou *extreme learning machine (ELM)*, organiza sua aprendizagem com base exclusivamente no *feedforward*, funcionando como aproximador universal (CAO; LIN; HUANG, 2012); desta maneira, a informação é transportada em uma única direção, da entrada para saída. A MLP, por exemplo, apresenta, além desse caminho natural dos dados, o *backward*, em que a informação ao chegar na saída é realimentada na entrada para atualização da própria rede, modificando os pesos sinápticos até que estes proporcionem uma saída próxima do desejado para os estímulos de treinamento (HAYKIN, 2011).

As vantagens da rede ELM sobre as demais paira justamente em seu “treinamento”: enquanto RNAs, em geral, necessitam de milhares de épocas com várias iterações cada, a ELM tem seu aprendizado em uma única etapa não recursiva. Segundo Huang, Wang e Lan (2011), a essência da rede compreende exatamente o uso de pesos sinápticos aleatórios na camada oculta, o que resulta em um modelo linear para os da camada de saída, determinados analiticamente a partir de solução do método dos mínimos quadrados.

Por outro lado, a principal limitação da ELM também deriva do uso de pesos aleatórios na camada escondida, uma vez que estes podem não oferecer um conjunto ótimo de operação da rede, em vista que não são atualizados durante o “treinamento”. Outro entrave advém da determinação do número de neurônios ocultos, pelo dilema de encontrar o meio termo entre a compensação da inicialização aleatória destes e a possibilidade de diminuir-los em quantidade sem afetar a eficácia do aprendizado. Essa resposta, em geral, é obtida por tentativa e erro (YANG; WANG; YUAN, 2012).

MATERIAIS E MÉTODOS

A rede ELM desse estudo foi arquitetada estruturalmente dentro do ambiente de programação *MATLAB*, com algoritmo *forward*. A função aproximada é originalmente apresentada em (1), uma combinação de senóide e exponencial; apresenta múltiplos máximos e mínimos locais dentro do intervalo escolhido para a aproximação, de zero a duas vezes o valor de π .

$$f(x) = 2\cos(3x) \cdot \text{sen}(0,25x) + 0.15(x - 3)^2 \quad (1)$$

Caracteristicamente, a função não é de simples aproximação por conta de seu comportamento senoidal não elementar. Para treinamento, etapa não-recursiva, entretanto, acrescentou-se ainda sinal de erro $e(x)$, como pode ser visto em (2); este termo representa um ruído branco gaussiano com amplitude da ordem de 10% da função original, que, em termos experimentais, é a modelagem de qualquer distúrbio adquirido durante uma aquisição de dados em um sistema prático.

$$f(x) = 2\cos(3x) \cdot \sin(0,25x) + 0.15(x - 3)^2 + e(x) \quad (2)$$

A rede foi estruturada com apenas uma camada escondida, segundo parâmetros usuais ocorre para a aproximação de função; nesta camada, testou-se a rede para cinco, dez, vinte e cinquenta neurônios. A modelagem do neurônio da rede inclui função de ativação não linear, com amplitude de -1 a 1 — essa função pode ter um padrão também de zero a um, mas se adota neste trabalho o primeiro intervalo por uma necessidade de se trabalhar com a região negativa da função aproximada. Para que a não linearidade da função de ativação seja essencialmente suave, adota-se neste estudo a função tangente hiperbólica, conforme (3).

$$f(x) = \text{tghyp}(x) = \frac{1 - e^{-2ax}}{1 + e^{-2ax}} \quad (3)$$

Para obtenção de dados, deliberou-se o estudo com base em repetições do treinamento e teste da rede a fim mais relevantes dentro do estudo, garantindo maior precisão e confiabilidade nos resultados. A quantidade de repetições utilizadas foi de 1000 vezes, com valores estatísticos discutidos na sessão seguinte.

RESULTADOS E DISCUSSÃO

A Figura 1 apresenta diversas curvas interessantes para avaliação do comportamento de saída da rede em comparação com a saída que ela deveria apresentar. A função original (curva tracejada em vermelho) apresenta a expressão em (1), enquanto que a função ruidosa representa esta mesma expressão acrescida de ruído gaussiano branco, conforme (2); a aproximação média é a média realizada entre as 1000 curvas obtidas para cada uma das repetições do par treino-teste da rede e a zona de aproximação corresponde a região de acomodação da saída da rede para todas as repetições, ou seja, é o intervalo que contém todas as possíveis soluções encontradas pela rede dentro das repetições realizadas.

Nitidamente, com o aumento da quantidade de neurônios na camada escondida, a aproximação realizada pela rede se torna mais acurada: mesmo com a presença do ruído, a saída média da rede com 50 neurônios corresponde praticamente à própria função original, conforme Figura 1f. No lado oposto, na Figura 1a, observa-se que a curva média obtida das saídas para 5 neurônios não consegue aprender corretamente a dinâmica da função, uma vez que, até se consegue acompanhar o crescimento médio da função, no entanto, os pontos de mínimo e máximos não são alcançados.

Com quinze neurônios escondidos, a aproximação média na saída da rede dentro das 1000 repetições já praticamente corresponde à função original, conforme Figura 1c. No entanto, a zona de aproximação permite observar que a quantidade de neurônios ainda não é satisfatória, pois os pesos sinápticos da camada oculta, inicializados aleatoriamente, ainda apresentam muita dependência de sua inicialização para apresentar um resultado adequado: a aleatoriedade pode gerar uma boa aproximação, ou uma com desvios não toleráveis.

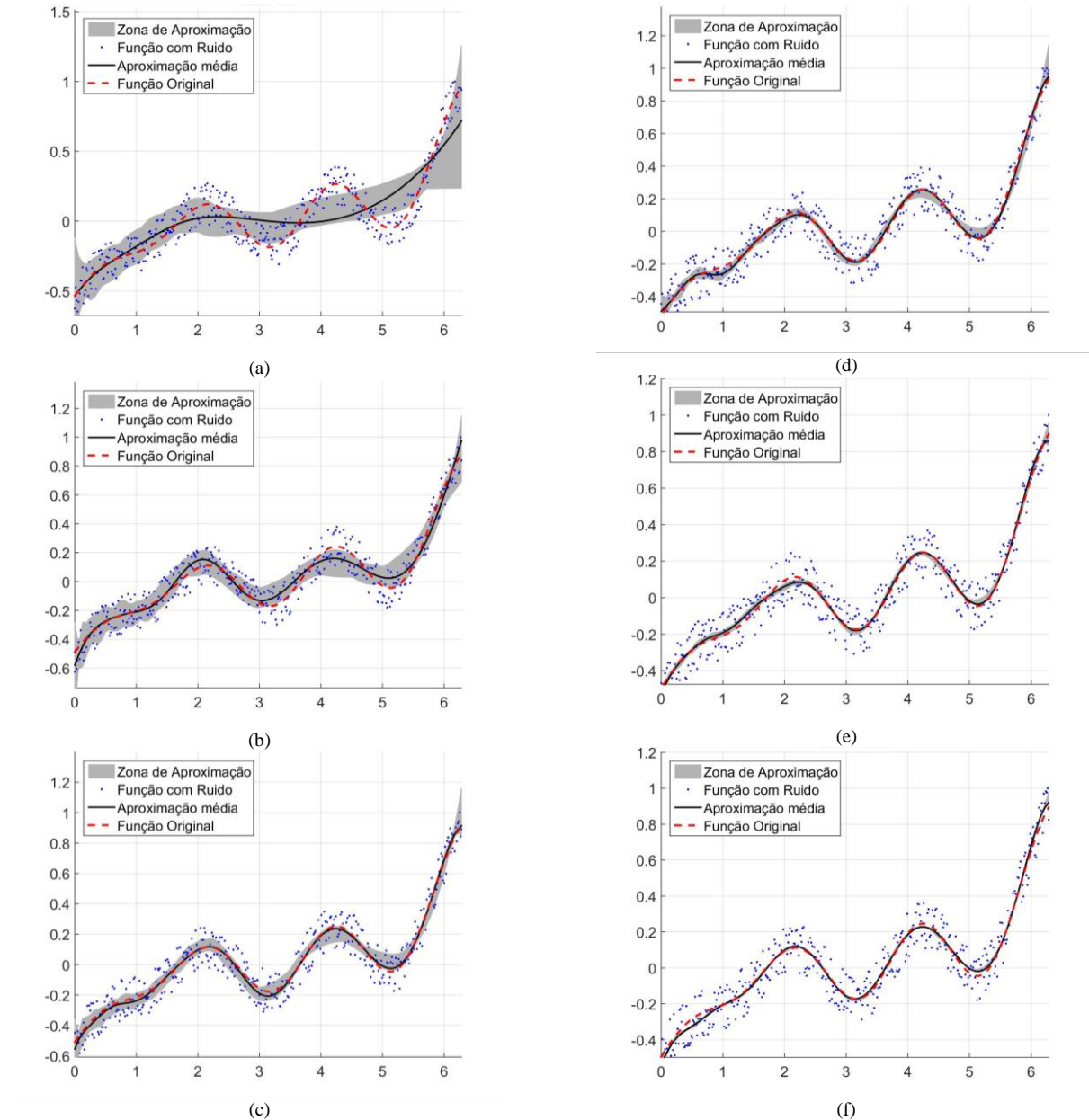
Por fim, a aproximação média da saída é satisfatória e a zona de aproximação bem diminuta para 30 neurônios (Figura 1e); a resposta ideal, entretanto, entre os valores de neurônios testados, é 50. Caso se estivesse realizando um estudo de dinâmica de um sistema qualquer em que se requeresse uma otimização da quantidade de neurônios na camada oculta, o valor exato ideal estaria entre 30 e 50 e poderia ser obtido por meio de tentativa e erro.

A Tabela 1 apresenta o erro quadrático médio e o erro absoluto médio dentro das 1000 repetições para o teste da rede com as diversas quantidades de neurônios testados. Os dados elencados confirmam o que é apresentado na Figura 1: quanto maior o número de neurônios escondidos, menor o erro. Percebe-se ainda uma tendência de saturação da diminuição do erro, em que o aumento de neurônios já não representaria uma diminuição significativa do erro, porém, aumentaria o tempo de rodagem da rede.

Tabela 1. Distribuição das classes do potencial pedológico da cultura da cana de açúcar

Número de neurônios	5	10	15	20	30	50
Erro quadrático médio	0,0523	0,0509	0,0484	0,0476	0,0475	0,0448
Erro absoluto médio	0,2417	0,2338	0,2293	0,2278	0,2208	0,2190

Figura 1. Curvas da função original, da com ruído gaussiano branco, da aproximação média pela rede e da zona de pertinência das saídas da rede para (a) 5, (b) 10, (c) 15, (d) 20, (e) 30 e (f) 50 neurônios na camada escondida.



CONCLUSÕES

A Máquina de aprendizagem extrema é um método bastante satisfatório para o estudo de aproximação de função, apresentando uma modelagem da função estudada satisfatória com poucos desvios quando configurada com os parâmetros adequados. O aumento de número de neurônios na camada escondida evidencia-se como fator determinante em uma boa aproximação média.

Pelas características avaliadas, um possível futuro estudo pode incluir a modelagem de sistemas dinâmicos com base no algoritmo já implementado, uma vez que se utilizou já previamente de ruído gaussiano branco em simulação a uma aquisição de dados real em que a obtenção de erro aglutinado ao dado verdadeiro é indissociável. O critério de quantidade de neurônios necessários para a aplicação faz-se necessário para qualquer nova aplicação, principalmente por tentativa e erro.

Por fim, a rede ELM apresenta excelente eficiência em seu desempenho: apesar de terem sido necessárias 1000 repetições de treino-teste do algoritmo elaborado para resultados estatisticamente

mais confiáveis, não foi necessário processamento computacional extenuante como ocorre para outras redes. Isso se deve em virtude da ausência de uma etapa de repetidas iterações de treino, característica marcante desta RNA.

AGRADECIMENTOS

Ao Programa de Educação Tutorial por propiciar ambiente saudável de pesquisa e pela concessão de bolsa aos dois primeiros autores.

REFERÊNCIAS

- Braga, A. P.; Ferreira, A. C. P. L.; Ludermir, T. B. Redes neurais artificiais: teoria e aplicações. LTC Editora, 2007.
- Cao, J.; Lin, Z.; Huang, G.-B. Self-adaptive evolutionary extreme learning machine. *Neural processing letters*, Springer, v. 36, n. 3, p. 285–305, 2012.
- Huang, G.-B.; Wang, D. H.; Lan, Y. Extreme learning machines: a survey. *International Journal of Machine Learning and Cybernetics*, Springer, v. 2, n. 2, p. 107–122, 2011.
- Louzada, Helder Araujo; Siravenha, Ana Carolina Quintão; Pelaez Evaldo Gonçalves. Utilização de redes neurais do tipo extreme learning machine na classificação da cobertura de solo do Município de Novo Progresso-PA. *Anais XVII Simpósio Brasileiro de Sensoriamento Remoto - SBSR*, João Pessoa-PB, Brasil, 25 a 29 de abril de 2015, INPE.
- Haykin, Simon. *Redes neurais: princípios e prática*. 2. ed. trad. Paulo Martins Engel. Porto Alegre: Bookman, 2001. pp. 27-282, 483-524.
- Negnevitsky, Michael. *Artificial intelligence: a guide to intelligent systems*. 2. Ed. King's Lynn: Pearson, 2007. pp. 175-185.
- Souza, L. G. M. *Proposição e avaliação de algoritmos de filtragem adaptativa baseados na rede de kohonen*. Fortaleza: UFCE, 2005.
- Souza, L. G. M. *Modelos lineares locais para identificação de sistemas dinâmicos usando redes neurais competitivas*. Fortaleza: UFCE, 2012.
- Von Zuben, F. J. *Modelos paramétricos e não-paramétricos de redes neurais artificiais e aplicações*. Campinas: FEE-UNICAMP, 1996.