

Predição do comportamento de manchas solares utilizando abordagens de redes neurais e técnicas de aprendizado de máquina

Sunspot behaviour prediction using neural network approaches and machine learning techniques

RESUMO

Sthefanie Monica Premebida
sthefanie@alunos.utfpr.edu.br
Universidade Tecnológica Federal do Paraná, Ponta Grossa, Paraná, Brasil

Marcella Scoczynski Martins
marcella@utfpr.edu.br
Universidade Tecnológica Federal do Paraná, Ponta Grossa, Paraná, Brasil

Virginia Helena Varotto Baroncini
virginia@utfpr.edu.br
Universidade Tecnológica Federal do Paraná, Ponta Grossa, Paraná, Brasil

O estudo da atividade solar é de grande interesse para o reconhecimento da sua influência na Terra. Um grande passo na astronomia é a previsão da atividade solar, permitindo maior preparo para estudo e reconhecimento de futuros eventos solares e terrestres. Na pesquisa, foi utilizado redes neurais e técnicas de *Machine Learning* para predição do número de manchas solares baseando-se na atividade solar registrada entre 1818 e 2019. Os dados referentes à atividade solar foram retirados do site do *Solar Influences Data analysis Center (SIDC)* e *Sunspot Index and Long-term Solar Observations (SILSO)*. Os resultados obtidos dos experimentos realizados mostram um desempenho competitivo desta abordagem na predição de manchas solares.

PALAVRAS-CHAVE: Manchas Solares. Redes Neurais. Aprendizado de Máquina. RBF. ELM.

Recebido: 19 ago. 2019.

Aprovado: 01 out. 2019.

Direito autoral: Este trabalho está licenciado sob os termos da Licença Creative Commons-Atribuição 4.0 Internacional.



ABSTRACT

THE STUDY OF SOLAR ACTIVITY IS OF GREAT INTERESTING OF THE INSTERESTING FOR THE RECOGNITION OF ITS INFLUENCE ON THE EARTH. A GREAT STEP IN ASTRONOMY IS THE PREDICTION OF SOLAR ACTIVITY, ALLOWING BETTER PREPARATION FOR STUDY AND RECOGNITION OF FUTURE SOLAR AND TERRESTRIAL EVENTS. IN THIS RESEARCH WAS USED NEURAL NETWORKS AND MACHINE LEARNING TECHNIQUES TO PREDICT SUNSPOT NUMBERS BASED ON SOLAR ACTIVITY RECORDED BETWEEN 1818 AND 2019. SOLAR ACTIVITY DATA WERE TAKEN FROM THE SOLAR INFLUENCES DATA ANALYSIS CENTER (SIDC) WEBSITE AND SUNSPOT INDEX AND LONG-TERM SOLAR OBSERVATIONS (SILSO). RESULTS SHOW A HIGH POTENTIAL OF THIS PROCESSING THAT BECOME A COMPETITIVE APPROACH FOR THE SUNSPOTS PREDICTION.

KEYWORDS: Sunspots. Neural Networks. Machine Learning. RBF. ELM.

INTRODUÇÃO

O Sol apresenta um ciclo de 11,2 anos, e após isso, todo o sol inverte sua polaridade magnética geral: o polo norte vira o polo sul e vice-versa [MALUF 2003]. Portanto, um ciclo magnético solar completo leva cerca de 22 anos, sendo conhecido como ciclo de Hale, mas o comportamento varia conforme a variação da atividade.

De acordo com Hathaway (1994), a observação de manchas solares e atividade solar a partir do século XVII mostrou que o número de manchas e a área que elas cobrem cresce rapidamente, de um mínimo (perto de zero) para o máximo (3 a 4 anos depois de atingir o mínimo). No entanto, o declínio máximo para o mínimo é mais lento. Este crescimento assimétrico e declínio exibe variações substanciais de um ciclo para o outro. Entre 1645 e 1715 ocorreu o maior mínimo solar da história, o mínimo Maunder, quando os ciclos de 11 anos foram interrompidos. É por isso que os ciclos solares foram numerados desde 1755 e 24 ciclos foram contados até o ano de 2019.

Atualmente, a maior referência para medir a atividade solar é o número de manchas solares presente no sol a qualquer momento. Desde 1981 a análise das imagens fornecidas por satélites e observatórios solares, para monitoramento automatizado, estão sendo realizadas pelo *Solar Influences Data analysis Center (SIDC)*. O SIDC produz mensalmente, há 38 anos, a relação internacional de manchas solares, R_i , que é calculado estatisticamente com os dados de todos os colaboradores e de observações disponíveis, as contribuições são de observadores profissionais e amadores de todo o mundo, especialmente da Europa Ocidental, com o número de Wolf. Para equalizar os dados para encontrar um R_i consistente, é usado um coeficiente de redução pessoal k , que é o fator da escala entre a estação individual e a média geral da rede [Clette et al. 2007].

Existem diversas previsões sendo feitas para o ciclo 25, de toda atividade solar já registrada, onde as previsões do SIDC são usadas como referência. Existem 3 métodos de previsão: curvas padrão, método combinado e McNish & Lincoln. De acordo com o SIDC, todos os métodos são baseados puramente na atividade solar, tornando-se pouco confiáveis no final de cada ciclo e durante os mínimos, porém os métodos têm uma boa performance durante o meio de cada ciclo.

Nesse trabalho propomos uma abordagem que faz a previsão do número de manchas solares utilizando redes neurais. Para isso, utilizamos técnicas de *machine learning* e redes neurais distintas, com variação numérica de neurônios. No total, 73565 dias foram analisados, todos registrados pelo SIDC.

MATERIAL E MÉTODOS

Manchas solares são recorrentes, elas aparecem e desaparecem em questão de dias ou semanas. Uma das maneiras para análise e tentativa de achar possíveis padrões de aparecimento é utilizando um banco de dados extenso, porém, o tempo de processamento será tão grande quanto o tamanho do banco. Para obtenção de resultados de uma maneira mais rápida, algumas ferramentas de computação e *machine learning* podem ser utilizadas, no nosso trabalho optamos por usar as redes *Radial Basis Function (RBF)* e a *Extreme Learning Machine (ELM)*.

Radial Basis Function (RBF), é uma rede neural linear que utiliza a técnica das redes supervisionadas (onde a rede aprende com um banco de dados pré-definido com as entradas e saídas, todas conhecidas). Nesse tipo de arquitetura, existe apenas uma camada intermediária, onde as funções de ativação são de alta dimensionalidade [Silva e Flauzino 2010]. A RBF pode ser usada como uma função básica para modelos de regressão não lineares (lineares ou não lineares) e pode ser utilizada como uma função de ativação de qualquer rede multicamadas [Von Zuben e Attux 2001], como a *Multilayer Perceptron* (MLP). A função não linear mais utilizada em uma rede é a Gaussiana, onde existe uma aproximação de comprimentos e centroides de cada neurônio, enquanto na camada de saída é utilizada uma função de ativação linear.

Extreme Learning Machine (ELM) é uma rede de avanço que pode ser utilizada para classificação, regressão, aproximação e muitas outras aplicações possíveis, se adaptando a vários tipos de pesquisa. A rede ELM é similar à rede MLP, mas com algumas diferenças, como a camada intermediária, a rede ELM não tem treinamento da sua camada intermediária, ou seja, os pesos sinápticos dessa camada são aleatórios, os pesos que possuem normalização são os pesos da camada de saída, portanto, a quantidade de processamento para o treinamento é muito menor. Como dito anteriormente, a ELM é parecida com a rede MLP, e os números de entrada e de saída variam conforme a necessidade de cada aplicação. Alguns autores consideram a ELM como uma nova forma de aprendizado da MLP [Huang et al. 2011].

Nessa pesquisa, foram analisados os dados disponíveis no SICD/SILSO para predição do número de manchas, para isso, utilizamos as redes citadas a cima para fazer essa estimativa de número de manchas.

O primeiro passo para isso é organizar os dados que serão utilizados pela inteligência artificial. Esses dados são divididos entre entradas e saídas e divididos em 3 partes. As proporções escolhidas foram:

- Treinamento – 70%
- Validação – 15%
- Testes – 15 %

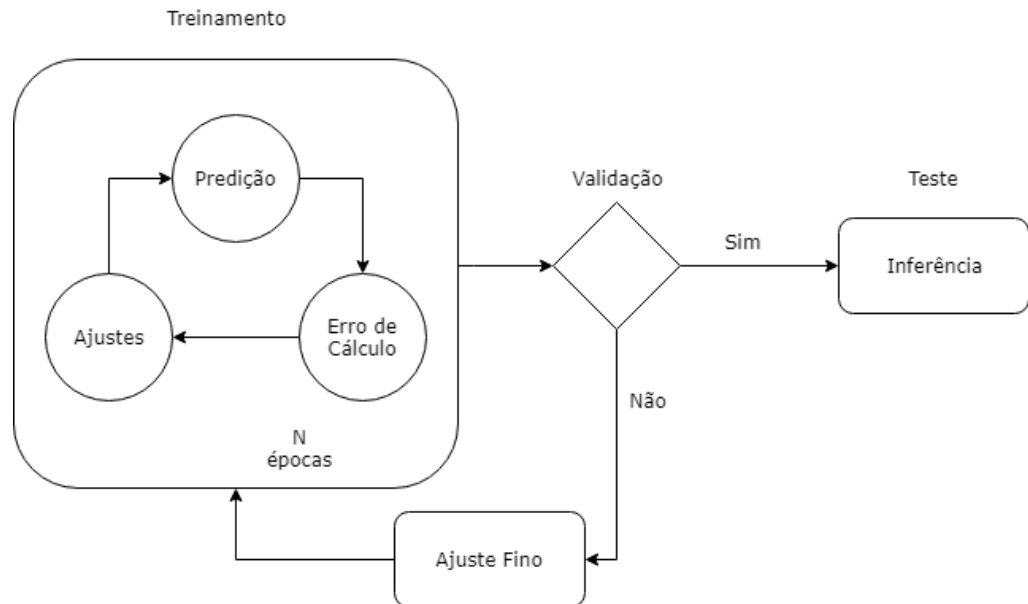
O processo de aprendizado ocorre através do número definido de épocas, onde o conjunto de treinamento é submetido a um loop de adivinhação dos resultados e tentativa de minimizar os erros. Durante cada época, as redes tentam imitar exatamente o mesmo comportamento do treinamento conjunto.

Após cada período ou n números de épocas, o conjunto de validação é usado para verificar se os resultados adivinhados são satisfatórios, e para analisar o comportamento da rede quando submetida para um novo conjunto de dados. Se esta etapa não for executada, a rede treinada responderá perfeitamente ao analisar os dados que já temos, mas perderá desempenho ao tentar prever com base em novos dados, que é o principal objetivo.

Dependendo dos resultados, a rede pode ser declarada treinada e pronta para executar previsões sobre o conjunto de testes ou passar por um ajuste fino e um novo ciclo de aprendizado e validação é iniciado.

Após o conjunto de testes ter sido submetido às redes agora treinadas (inferência), os resultados esperados e dados são comparados e o erro quadrático médio é usado como indicativo de desempenho. O processo de aprendizagem e inferência pode ser visto na Figura 1.

Figura 1 – Fluxograma do processamento de dados.



Fonte: Autoria própria (2019).

EXPERIMENTOS E RESULTADOS

No total, foram analisados 73565 dias, começando em 1º de janeiro de 1818 até 31 de maio de 2019. O formato utilizado para dados segue ano, mês, dia, data em fração do ano, total diário, número de manchas solares, desvio padrão diário dos números de manchas solares de entrada de cada estação, número de observações utilizadas para calcular o valor diário, definitivo/provisório indicador, nenhuma modificação foi feita.

Foi investigado a rede neural para prever o desempenho calculando o erro quadrático médio para todos os dias, dado pela Eq. (1) onde **nd** representa o número de dias, **R_{sv}** o valor real da quantidade de pontos diários e **P_{sv}** o valor da quantidade de pontos diários previstos.

$$MSE = \frac{1}{2*nd} * \sum_{i=0}^{nd} (R_{sv} - P_{sv})^2 \quad (1)$$

O processo de treinamento foi executado em 3 redes neurais diferentes com os mesmos recursos. As tabelas 1 e 2 apresentam os melhores e os piores valores de erro quadrático médio (*Mean Square Error - MSE*) para o processo de teste de diferentes quantidades de neurônios.

A Tabela 1, apresenta o erro quadrático médio (MSE) para a rede RBF com ELM. Nota-se que assim que a quantidade de neurônios aumenta, MSE diminui, mas o tempo de processamento aumenta.

Tabela 1 – ELM – RBF MSE.

Quantidade de neurônios	Melhor	Pior
500	0,0213	0,0245
1000	0,0109	0,0127
1500	0,0096	0,0103
2000	0,0078	0,0084

Fonte: Autoria própria (2019).

A Tabela 2 mostra o MSE para a ELM estruturada apenas com uma camada. Diferente dos resultados da Tabela 1, os resultados do MSE não diminuem conforme a quantidade de neurônios aumenta, fica quase inalterado, com uma amplitude de 0,0014 e não existe um ganho quando a quantidade de neurônios aumenta.

Tabela 2 – ELM 1 Camadas MSE.

Quantidade de neurônios	Melhor	Pior
500	0,0134	0,0137
1000	0,0129	0,0129
1500	0,0126	0,0126
2000	0,0123	0,0124

Fonte: Autoria própria (2019).

Utilizando uma ELM com 2 camadas, também obtivemos resultados semelhantes ao da Tabela 2, onde a variação de neurônios nas camadas não altera significativamente o valor de MSE, mas diminui o desempenho de previsão.

A título de comparação, foi considerado um trabalho apresentado por [Safiullin et al. 2018], onde os autores trabalharam com a mesma estrutura de dados do SIDC/SILSO, mas diferente deste, eles utilizaram um número total de manchas solares durante 13 meses e ao invés do MSE, foi utilizado o RMSE, que é a raiz quadrada do MSE, a fim de investigar o desempenho da rede neural. O melhor RMSE de [Safiullin et al. 2018] usando os dados SIDC sem adaptação é de 1,59, em comparação com o melhor RMSE presente aqui que é 0,0883, adaptado do valor de MSE 0,0078 presente na Tabela 1, onde temos a junção da rede RBF e ELM (a adaptação é obtida extraindo a raiz quadrada do MSE). Como pode-se notar, o presente trabalho tem maior desempenho utilizando o RMSE como precisão de previsão. A comparação é apresentada na Tabela 3.

Tabela 3 – ELM – RBF MSE.

RMSE	Essa pesquisa	Safiullin
	0,0883	1,5900

Fonte: Autoria própria (2019).

CONCLUSÃO

Esse trabalho investigou a utilização de redes neurais e técnicas de *machine learning* para prever o comportamento das manchas solares na superfície solar. Os experimentos abordam dados de 73565 dias registrados pelo *Solar Influences Data analysis Center* (SIDC). Foram analisados 3 tipos diferentes de redes neurais com as mesmas características e após isso, foi feita a análise de desempenho de acordo com o erro quadrático médio (MSE).

Além disso, foi comparada a ELM com método RBF com um artigo relacionado presente na literatura, que utiliza a raiz do erro quadrático médio (RMSE) para calcular o desempenho de sua rede. O resultado mostra que usar a ELM com método RBF fornece um resultado competitivo de acordo com o resultado do RMSE. No futuro, planejamos identificar outras características relevantes do Sol que possam ser previstas e classificadas, melhorando a pesquisa nessa área espacial.

AGRADECIMENTOS

Agradeço primeiramente à Deus, meus colegas de pesquisa, Thiago F. O. de Camargo, Vinícios R. Soares e Denise Pechebovicz e às minhas professoras, Marcella S. Martins e Virgínia H. V. Baroncini por todo apoio e cuidado durante as pesquisas. Meus colegas e professores do LICON, laboratório onde a pesquisa foi desenvolvida, também tem meus mais sinceros agradecimentos.

REFERÊNCIAS

CLETTE, Frédéric et al. From the Wolf number to the International Sunspot Index: 25 years of SIDC. *Advances in Space Research*, v. 40, n. 7, p. 919-928, 2007.

HUANG, Guang-Bin et al. Extreme learning machine for regression and multiclass classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, v. 42, n. 2, p. 513-529, 2011.

HATHAWAY, D. H. The solar dynamo. *NASA STI/Recon Technical Report N*, 96. 1994.

ECHER, Ezequiel et al. O número de manchas solares, índice da atividade do sol. *Revista Brasileira de Ensino de Física*, v. 25, n. 2, p. 157-163, 2003.

SAFIULLIN, Nikolai; PORSHNEV, Sergey; KLEEORIN, Nathan. Monthly sunspot numbers forecast with artificial neural network combined with dynamo model: comparison with modern methods. In: 2018 Ural Symposium on Biomedical Engineering, Radioelectronics and Information Technology (USBREIT). IEEE, 2018. p. 199-202.

DA SILVA, I. N.; SPATTI, D.; FLAUZINO, R. *Redes Neurais Artificiais para engenharia e ciências aplicadas: curso prático*, Artliber Editora Ltda, São Paulo, SP, Brasil. ISBN978-85-88098-53-4, 2010.

VON ZUBEN, IA353—Profs Fernando J.; ATTUX, Romis RF. *Redes Neurais com Funções de Ativação de Base Radial*. DCA/FEEC/Unicamp, 2001.