

Previsão de manchas solares usando abordagens de redes neurais e técnicas de aprendizado de máquina

Sunspot behavior prediction using neural network approaches and machine learning techniques

RESUMO

Denise Pechebovicz
pechebovicz@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

O estudo da atividade solar é de grande interesse para o reconhecimento de sua influência sobre a terra. Um grande passo na astronomia é a previsão da atividade solar, permitindo melhor preparação para estudos e reconhecimento de futuros eventos solares e terrestres. Em nossa pesquisa, utilizamos redes neurais e técnicas de aprendizagem de máquina para prever número de manchas solares com base na atividade solar gravada entre 1818 e 2019. Os dados da atividade solar foram obtidos no site *Solar Influences Data analysis Center* (SIDC) e *Sunspot Index and Long-term Solar Observations* (SILSO). Os resultados mostram um elevado potencial de processamento, o que se torna uma proposta competitiva para a previsão manchas solares.

PALAVRAS-CHAVE: Processamento de imagem. Rede neural de convolução. Classificação. Manchas Solares.

Recebido: 19 ago. 2019.

Aprovado: 01 out. 2019.

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



ABSTRACT

THE STUDY OF SOLAR ACTIVITY IS OF GREAT INTEREST 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 OUR RESEARCH WE 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: Image processing; convolutional neural network; classification; sunspots.

INTRODUÇÃO

É possível entender as manchas solares presentes na superfície solar como regiões possuintes de intensos campos magnéticos que bloqueiam a matéria convectiva abaixo da fotosfera, formadas por um processo de dínamo solar.

O ciclo solar magnético completo dura em média 22 anos, em que a cada 11,2 o sol inverte toda sua polaridade magnética geral, sendo conhecido como ciclo Hale, mas o comportamento varia com a variação da atividade solar. Segundo Hathaway (1998), observações de manchas solares e atividade solar a partir do século XVII mostraram que o número de manchas solares e a área que cobrem crescem rapidamente de um mínimo (próximo de zero) ao máximo (3 a 4 anos após atingir o mínimo). No entanto, a redução do máximo para o mínimo é mais lenta. Este crescimento assimétrico e declínio inibem variações substanciais de um ciclo para outro. Os ciclos solares foram numerados desde 1755 e 24 ciclos foram contados até o ano atual, 2019.

Atualmente, a maior referência para medir a atividade solar é o número de manchas solares presentes no sol a um determinado momento. Desde 1981 a análise de imagens provenientes de satélites e observatórios, por processos automatizados, é feito pelo *Solar Influences Data Analysis Centes* (SIDC). Esse centro possui 38 anos e produz mensalmente o *International Relative Sunspot Number*, R_i , que é calculado por através de todos os contribuintes e observações disponíveis, observadores profissionais e astrônomos amadores de vários países, especialmente no leste da Europa, com o *Wolf Number*. Em função de igualar os dados para obter um R_i razoável, usa-se o coeficiente de redução pessoal k , que é um fator de escala entre a estação individual e o total da rede.

Há inúmeras previsões sendo feitas para o ciclo 25, a partir de toda a atividade solar já registrada, onde as previsões do SIDC são usadas como referência. Existem três métodos de predições: Standard Curves, Combined method e McNish & Lincoln. Segundo o SIDC, todos os métodos baseados puramente na atividade solar passada, torna-se não confiável no final de cada ciclo e durante os mínimos. Mas este método possui boa performance no meio de cada ciclo.

Neste trabalho, propomos uma abordagem com o intuito de fazer predições do número de manchas solares usando inteligência artificial. Para este propósito, usamos técnicas de aprendizado de máquina e diferentes redes neurais com variação no número de neurônios, analisando, no total, 73565 dias registrados pelo SIDC.

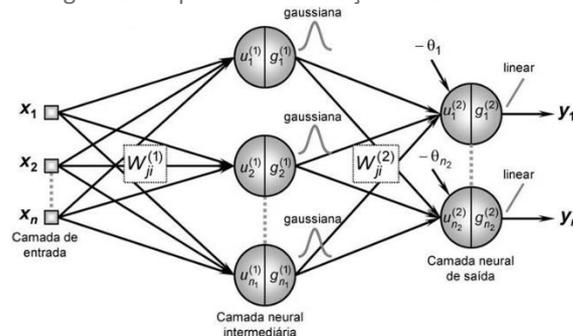
MATERIAL E MÉTODOS

As manchas solares são recorrentes, ou seja, aparecem e desaparecem em questão de dias ou semanas. Uma maneira de analisar e tentar encontrar possíveis padrões é usando um grande banco de dados para análise, mas essa alternativa utiliza de um longo tempo de treinamento de rede. Para resultados mais rápidos, utilizamos métodos computacionais e de aprendizado de máquina, que neste caso são utilizados Função de Base Radial (RBF) e Máquina de Aprendizado Extremo (ELM).

Função de Base Radial (RBF) é uma rede neural linear e tem a sua aprendizagem supervisionada (onde a rede aprende com o conjunto de dados já definido e as entradas e saídas são conhecidas). Neste tipo de arquitetura, existe apenas uma camada intermediária, onde as funções de ativação são de alta dimensionalidade. A RBF pode ser usada como uma função básica para modelos de regressão (linear ou não-linear), também como uma função de ativação de qualquer tipo de rede multicamadas, tal como Perceptron Multicamada (MLP). A

função não linear mais utilizada em uma rede é a Gaussiana, onde temos uma aproximação de centroides e comprimento de cada neurônio, enquanto na camada de saída, é utilizada uma função de ativação linear.

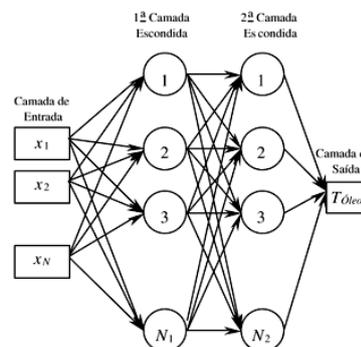
Figura 1: Arquitetura da Função de Base Radial



Fonte: Slides “RNA – Radial Basis Function” (2015).¹

Máquina de Aprendizagem Extrema (ELM) são redes neurais feedforward e pode ser utilizada para classificação, regressão, aproximação e tantas outras aplicações em diversas pesquisas. ELM é semelhante a um Perceptron Multicamada (MLP) (Fig. 2), onde a quantidade de entradas e saídas são variáveis de acordo com as necessidades ou aplicações. Porém possui algumas diferenças, como na camada intermédia, em que ELM executa ajustes na camada intermédia, isto é, os pesos sinápticos nesta camada são dados ao acaso, conseqüentemente, apenas os pesos da camada de saída são sintonizados, de modo que a quantidade de processamento para o treinamento é muito menor. Uma vez que a rede tem esta aleatoriedade, é classificada como uma máquina desorganizada.

Figura 2: Arquitetura Perceptron Multicamadas



Fonte: Sba Controle & Automação vol.13 no.3 Campinas (2002).²

Este trabalho apresenta uma abordagem para prever o número de manchas solares analisando os dados disponíveis no SIDC / SILSO. Foram aplicadas técnicas de aprendizado de máquina para encontrar um padrão, a fim de estimar o número de manchas solares. O primeiro passo é organizar os dados a serem usados pela inteligência artificial. Os dados primários foram divididos em entradas e saídas em que são divididos em três partes, treino, validação e teste.

¹ Disponível em: < <https://slideplayer.com.br/slide/3381271/> > Acesso em: 19 de agosto de 2019.

² Disponível em: < <http://dx.doi.org/10.1590/S0103-17592002000300005> > Acesso em: 19 de agosto de 2019.

As proporções escolhidas foram as seguintes: Treinamento - 70%; conjunto de validação - 15%; conjunto de testes - 15%. O processo de aprendizagem ocorre através de um número definido de épocas, onde o conjunto de treinamento é submetido a um loop de adivinhar os resultados, sugerindo os mesmos e minimizando os erros conforme os acertos.

Durante cada época, as redes tentam imitar exatamente o mesmo comportamento do conjunto de treinamento. Após esta ou inúmeras épocas, o conjunto de validação é usado para verificar se os resultados calculados são satisfatórios, e também analisar o comportamento da rede quando submetida a um novo conjunto de dados. Se esta etapa não for executada, a rede responderá perfeitamente ao analisar os dados treinados, porém perde desempenho ao tentar prever com base em novos dados, nosso principal objetivo.

Dependendo dos resultados, a rede pode ser declarada treinada e pronta para executar a previsão sobre o conjunto de dados de teste ou passar por um ajuste e um novo aprendizado e validação começa, como um ciclo. Após o conjunto de testes ter sido enviado para as redes agora treinadas (inferência), os dados e resultados esperados são comparados e usado o erro quadrático médio como um indicativo de desempenho.

RESULTADOS E DISCUSSÃO

No total, foram analisados 73565 dias, entre 1º de janeiro de 1818 e 31 de maio de 2019. O formato de dados segue ano, mês, dia, data em fração do ano, número total de manchas solares, desvio padrão diário do número de manchas solares de entrada de estações individuais, número de observações usadas para calcular o valor diário e indicador definitivo / provisório.

Nós investigamos o desempenho de previsão pela Rede Neural computando o erro quadrático médio (MSE) para todos os dias, que é dada pela Equação 1:

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

Onde nd representa o número de dias, R_{sv} o valor diário real da quantidade de manchas e P_{sv} o valor diário real previsto de quantidade de manchas. O processo de treinamento executa três redes neurais diferentes com os mesmos recursos. As Tabelas I, II e III apresentam os melhores e piores MSE para o processo de teste com diferentes quantidades de neurônios.

TABELA I: MSE para ELM+RBF

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 II apresenta o MSE para a Função de Base Radial com a Máquina de Aprendizado Extrema. Como já previsto com o aumento da quantidade de neurônios o MSE diminui, porém, o tempo de processamento também aumenta.

TABELA II: MSE para ELM com 1 camada

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: Aatoria Própria (2019).

A Tabela III mostra o MSE para o ELM estruturado com apenas uma camada oculta. Diferentemente dos resultados da Tabela I, o MSE não diminui significativamente com o aumento da quantidade de neurônios e a amplitude é de 0,0014.

Tabela III: MSE para ELM com 2 camadas

Quantidade de neurônios		Melhor	Pior
Camada 1	Camada 2		
500	500	0.049	0.0464
1000	500	0.0464	0.0469
1500	500	0.0462	0.0465
2000	500	0.0453	0.0460
500	1000	0.0446	0.0457
1000	1000	0.0454	0.0466
1500	1000	0.0457	0.0476
2000	1000	0.0459	0.0464
500	1500	0.0451	0.0454
1000	1500	0.0453	0.0457
1500	1500	0.0453	0.0458
2000	1500	0.0456	0.0475
500	2000	0.0445	0.0459
1000	2000	0.0457	0.0463
1500	2000	0.0457	0.0471
2000	2000	0.0454	0.0465

Fonte: Aatoria Própria (2019).

O mesmo fenômeno observado na Tabela II ocorre na Tabela III, significando que a variação do número de neurônios não afeta o MSE, mas diminui o desempenho de previsão porque a escala MSE é calculada por um fator de soma médio de 0,0032. Em geral, para as três abordagens, a proposta ELM com RBF alcançou os menores erros com 2000 neurônios na camada oculta. Para comparação, consideramos o trabalho apresentado por [Safiullin et al. 2018], onde os autores trabalharam com os mesmos dados de estrutura de SIDC / SILSO, mas diferentemente do nosso, eles usaram 13 meses alisando o número total mensal de manchas solares e a raiz do erro quadrático médio (RMSE) ao invés do MSE para investigar o desempenho da rede neural. Nesse caso obtiveram o melhor RMSE, usando dados SIDC sem adaptação, de 1.59. Nosso melhor resultado é MSE = 0,0078, ou o equivalente, RMSE = 0,0883 para o modelo ELM + RBF.

CONCLUSÃO

Este trabalho investigou redes neurais e técnicas de aprendizado de máquina para prever o comportamento de manchas solares na superfície solar. Nossos experimentos abordaram imagens registradas em 73565 dias pelo Centro de Análise de Dados de Influências Solares (SIDC).

Analisamos três redes neurais diferentes: Máquinas de Aprendizado Extremo, Função de Base Radial e uma proposta usando ambos os modos. Medimos o desempenho de acordo com o erro quadrático médio (MSE). Os resultados mostraram que o modelo híbrido poderia superar os outros. Além disso, estendemos nossa análise comparando o ELM + RBF com um artigo relacionado para avaliar a precisão da previsão. Os resultados mostraram que nosso método fornece desempenho competitivo.

No futuro, esperamos investigar mais técnicas para identificar outras características relevantes do sol a serem previstas e classificadas, melhorando a pesquisa nesta área.

AGRADECIMENTOS

Denise Pechebovicz agradece a Marcella Martins, Virginia Baroncini e Hugo Siqueira pela ajuda no desenvolvimento dessa pesquisa como também ajuda financeira e logística. Também agradece aos membros da equipe de iniciação científica, Sthefanie Premebida, Thiago Camargo e Vinícios Soares, pelo apoio e trabalho realizado.

REFERÊNCIAS

Hathaway, D. H. **The solar dynamo**. NASA Technical Report NASA-TM-111102, NAS 1.15:111102, 1994.

Clette, F.; Berghmans, D.; Vanlommel, P.; Van der Linden, R. A.; Koeckelenbergh, A.; Wauters, L. **From the wolf number to the international sunspot index: 25 years of sidc**, Advances in Space Research, vol. 40, no. 7, pp. 919–928, 2007.

Silva, I. N.; Spatti, D.; Flauzino, R. **Redes Neurais Artificiais para engenharia e ciências aplicadas: curso prático**, Artliber Editora Ltda, São Paulo, SP, Brasil, 2010.

Von Zuben, I. P. F. J.; Attux, R. R. **Redes Neurais com Funções de Ativação de Base Radial**. DCA/FEEC/Unicamp, 2001.

Echer, E. 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.