

Classificação de manchas solares utilizando técnicas de pré-processamento de imagem e técnicas de aprendizado profundo

Solar spots classification using pre-processing and deep learning image techniques

RESUMO

Técnicas de aprendizado de máquina e processamento de imagens têm sido aplicadas com sucesso em muitos campos de pesquisa. Astronomia e Astrofísica são algumas dessas áreas. Neste trabalho, aplicamos técnicas de aprendizado de máquina em uma nova abordagem para classificar e caracterizar manchas solares que aparecem na fotosfera solar que expressam campos magnéticos intensos, e esses campos magnéticos apresentam efeitos significativos na Terra. Em nossos experimentos, consideramos imagens de Heliosismism e Magnetic Imager (HMI) no formato IntensitygramFlat. Aplicamos técnicas de pré-processamento para reconhecer e contar os grupos de manchas solares para classificação adicional. Além disso, investigamos o desempenho da CNN AlexNet usando os dados como entrada em comparação com a Rede de Função de Base Radial (RBFN) usando diferentes níveis e combinando ambas as abordagens de redes. Os resultados mostram que quando a CNN usa o RBFN para identificar e classificar as manchas solares do processamento de imagens, seu desempenho é maior do que quando apenas a CNN é usada.

PALAVRAS-CHAVE: Processamento de imagens. Astronomia e Astrofísica. Redes neurais.

ABSTRACT

MACHINE LEARNING TECHNIQUES AND IMAGE PROCESSING HAVE BEEN SUCCESSFULLY APPLIED IN MANY RESEARCH FIELDS. ASTRONOMY AND ASTROPHYSICS ARE SOME OF THESE AREAS. IN THIS WORK, WE APPLY MACHINE LEARNING TECHNIQUES IN A NEW APPROACH TO CLASSIFY AND CHARACTERIZE SOLAR SPOTS WHICH APPEAR ON THE SOLAR PHOTOSPHERE WHICH EXPRESS INTENSE MAGNETIC FIELDS, AND THESE MAGNETIC FIELDS PRESENT SIGNIFICANT EFFECTS ON EARTH. IN OUR EXPERIMENTS WE CONSIDER IMAGES FROM HELIOSEISMIC AND MAGNETIC IMAGER(HMI) IN INTENSITYGRAMFLAT FORMAT. WE APPLY PRE-PROCESSING TECHNIQUES TO RECOGNIZE AND COUNT THE GROUPS OF SUNSPOTS FOR FURTHER CLASSIFICATION. BESIDES, WE INVESTIGATE THE PERFORMANCE OF THE CNN ALEXNET LAYER INPUT IN COMPARISON WITH THE RADIAL BASIS FUNCTION NETWORK (RBFN) USING DIFFERENT LEVELS AND COMBINING BOTH NETWORKS APPROACHES. THE RESULTS SHOW THAT WHEN THE CNN USES THE RBFN TO IDENTIFY AND CLASSIFY SUNSPOTS FROM IMAGE PROCESSING, ITS PERFORMANCE IS HIGHER THAN WHEN ONLY CNN IS USED.

KEYWORDS: Image Processing, Astronomy and Astrophysics, Neural Networks.

Vinícios Roberto Soares
viniciossoares@alunos.utfpr.edu.br
Universidade Tecnológica Federal do Paraná, Ponta Grossa, PR, Brasil

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

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

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.



INTRODUÇÃO

Erupções na superfície solar estão correlacionadas com manchas solares e há uma probabilidade de um evento ser avaliado com base na área, classe e tempo de vida do local (GIOVANELLI, 1939). Nesse sentido, Ejeções de Massa Coronal (CME) apresentam efeitos significativos na civilização da Terra (SISCOE, 2000). No entanto, a previsão de labaredas solares e esses efeitos ainda são difíceis de serem executados (SCHWENN, 2005). Ejeção de massa coronal é uma liberação de grande quantidade de plasma fechado com campo magnético da Corona Solar, labaredas solares consistem em um flash de luz, ocorre perto de um ponto solar e muitas vezes são seguidos por um CME.

Atualmente, a maioria dos benchmarks para medir a atividade solar considera o número de manchas solares presentes no sol a qualquer momento. Destacamos que é muito importante a anotação, contagem e a classificação das mesmas. No entanto, existem alguns estudos sobre o processamento de imagens diferentes para facilitar a classificação e contagem de manchas solares. Desde 1981 a análise de imagens fornecidas por satélites e observatórios do sol, para monitoramento automatizado, tem sido feita pelo Centro de Análise de Dados de Influências Solares (SIDC), que produz mensalmente o Número Internacional de Manchas Globais, R_i , calculado estatisticamente. Todos os contribuidores e observações disponíveis, usando o número de Wolf. Para equalizar os dados para encontrar um R_i consistente, é usado o coeficiente de redução pessoal k , que é a escala de fator entre a estação individual e a média geral da rede (CLETTE, 2007).

Este artigo tem como objetivo comparar a performance de redes neurais artificiais usando aprendizado profundo e aprendizado de máquina na tarefa de classificar manchas solares. Os principais objetivos aqui são identificar e classificar as manchas solares do processamento de imagens para serem mais exploradas para medir a atividade solar. Para tanto, utilizamos a análise de software de comunicação numérica e visualização de dados, MatLab.

Trabalhos similares abordaram as manchas solares em um contexto de processamento de imagens, especialmente usando a visão computacional, mas aqui, em nossa proposta, pretendemos aplicar técnicas de aprendizado de máquina que podem ser mais exploradas nas abordagens de processamento gráfico.

MATERIAIS E MÉTODOS

Atualmente, as manchas solares são as principais referências para determinar o nível de atividade solar. Além disso, as imagens captadas do sol são base de vários estudos para desenvolver teorias e melhor compreensão da estrela. A qualidade, o bom uso e a facilidade de observação das imagens são essenciais e, para melhorar essas características, técnicas de pré-processamento podem ser abordadas. Esta seção apresenta um plano de fundo para técnicas de pré-processamento e aprendizado de máquina, geralmente aplicadas à manipulação de imagens em um contexto geral.

Técnicas de pré-processamento - Ao manipular imagens, algumas técnicas devem ser aplicadas antes do processamento completo. Essas técnicas podem ajudar a fazer uma otimização de dados: preenchimento de nulos, tratamento de ruído, identificação ou remoção de valores discrepantes e remoção de inconsistências; integrar dados; processamento e redução de dados com valor particular para dados numéricos; normalização e agregação; discretização dos dados (J, H. A., & KAMBER, M., 2000). Listamos alguns métodos para realizar este processamento da seguinte maneira:

Ajuste de imagem (alongamento) => RGB para cinza=> Imagem para preto e branco=> Imagem aberta (crescimento) => Complemento de imagem=> Propriedades de região=> Recortar imagem=> Salvar imagem.

As técnicas de Aprendizado de Máquina podem ser usadas em várias situações automatizadas, pois podem produzir modelos rápida e automaticamente capazes de analisar dados maiores e mais complexos e entregar resultados mais rápidos e precisos, mesmo em larga escala (SMOLA, 2008). Aprendizado de Máquina é uma parte da Inteligência Artificial (IA) e apresenta quatro principais grupos de abordagens: aprendizado supervisionado, aprendizado não supervisionado, aprendizado semi-supervisionado e aprendizado por reforço. Nesse trabalho foi aplicado aprendizado supervisionado usando redes neurais de convolução (CNN) e Rede de Funções de Base Radial (RBFN).

A CNN escolhida foi a Alexnet, ganhadora da competição IMAGENET em 2012, uma CNN utiliza uma imagem como entrada e através de aprendizado profundo em suas camadas de convolução retira os parâmetros utilizados na classificação das categorias.

RBFN's utilizam funções de base radial em seus neurônios, fazendo com que a disposição espacial das entradas também se torne um fator relevante na classificação (média da distância ou a distância em relação a um ponto).

Este trabalho apresenta uma abordagem para classificar as imagens do sol de acordo com as manchas e algumas características especiais. Abordamos um conjunto de dados retirado do Heliosismism and Magnetic Imager (IHM), que processa imagens do Solar Dynamics Observatory (SDO).

A classificação é realizada de acordo com duas etapas: Primeira etapa: pré-processamento; Segunda etapa: fase de treinamento, utilizando CNN e a RBFN.

Primeiramente, as imagens foram submetidas a técnicas de pré-processamento, como alongamento, limiar, análise de propriedades de objeto e recorte. Após o pré-processamento, a fase de treinamento usa dois grupos de pontos: tipo O positivo e tipo O negativo. O nome dado as classes é arbitrário, definido para simplificar a análise: é uma classe basicamente caracterizada por quase apenas penumbra. O critério usado aqui é baseado na quantidade de pixels pretos para cada subfigura.

O método utilizado para reformular a AlexNet para o nosso propósito é transferir aprendizado, utilizar o conhecimento já adquirido pela rede em classificar imagens do cotidiano e transferir esse conhecimento para uma aplicação específica como o caso da classificação de manchas solares.

RESULTADOS E DISCUSSÃO

No total, imagens de 96 no formato HMI Intensitygram Flat foram analisadas, das quais 920 subfiguras foram geradas, 138 para validação, outros 138 para testes e 644 para a fase de treinamento. Exceto pela camada de dados do AlexNet, a quantidade de imagens tinha que ser menor que o total, essa redução foi necessária porque a nossa máquina de processamento não pôde suportar, portanto, o total de imagens foi reduzido para 400. Usamos imagens do Intensitygram Flat Orange 4K (disponíveis em [HTTP://HMI.STANFORD.EDU/](http://hmi.stanford.edu/)). Este formato permite a fácil manipulação para contagem, identificação e classificação de manchas solares a partir do processamento de imagens.

Nós investigamos o desempenho de redes neurais calculando três coeficientes, precisão, recordação e precisão. Tudo isso depende de alguns componentes, resultados da previsão de redes. Verdadeiro positivo (TP), este componente é dado quando há tipo O positivo na subfigura e o NN o detecta.

Falso positivo (FP), quando há O negativo na subfigura e o NN é classificado como O como positivo. Verdadeiro negativo (TN) é dado pela detecção de O negativo corretamente. Falso negativo (FN), este componente é dado quando o NN classifica uma subfigura contendo O positivo como O negativo.

Os critérios utilizados para análise foram os seguintes:

- a) Recall – Capacidade das redes em classificar corretamente;
- b) Precision – A proporção de verdadeiros positivos corretamente classificados;
- c) Accuracy – Porcentagem de classificações corretas.

Os resultados obtidos pela classificação usando diretamente imagens e aprendizado profundo podem ser observados na Tabela 1, mostrando o desempenho da Alexnet e os resultados obtidos utilizando aprendizado de máquina, com os parâmetros já extraídos na Tabela 2.

Tabela 1 – Análise de Performance Alexnet

Recall	0,8710
Precision	0,7714
Accuracy	0,9130

Fonte: Autoria Própria

Tabela 2 – Análise Comparativa

Rede	Recall	Precision	Accuracy
FC6-SoftMax	1	0,939	0,9333
FC6- Hyperbolic Tangent	1	0,9482	0,9619
FC8 - SoftMax	0,9732	0,8925	0,9167
FC8- Hyperbolic Tangent	1	0,9638	0,9381
Data - SoftMax	0	0	0,7452
Data - Hyperbolic Tangent	1	0,6063	0,4929

Fonte: Autoria Própria

Comparando o desempenho da RBFN de acordo com a entrada da camada da AlexNet, percebe-se que o melhor resultado obtido é com a camada 8 totalmente conectada, com método de tangente hiperbólica, e o pior resultado é dado pela camada totalmente conectada com o método softmax. Os resultados alcançados são maiores que o da AlexNet, isso pode ser explicado pela combinação das duas redes, AlexNet processa imagens e a Radial Basis Function Network classificam apenas dados processados. O desempenho da RBFN isolada está longe de poder ser comparado com a AlexNet, tabelas 1 e 2. A forma de desenvolvimento da RBFN não permite o processamento de um grande banco de dados, mas com um banco de dados pré-processado, como as 6 e 8 camadas totalmente conectadas, o RBFN é capaz de executar uma classificação razoável.

CONCLUSÃO

Este trabalho investigou um método para identificar e classificar manchas usando técnicas de processamento de imagens. Este método consiste em duas etapas: pré-processamento de imagem e fase de treinamento usando tanto a rede neural convolucional (CNN) quanto a Rede de Função de Base Radial (RBFN).

As imagens abordadas foram tiradas do Heliosismism and Magnetic Imager (IHM), do Solar Dynamics Observatory (SDO).

Um total de 96 imagens foi analisado, das quais 920 subfiguras foram geradas na fase de pré-processamento, separadas em 138 para testes, 138 para validação e 644 para treinamento. Analisamos o desempenho da rede de acordo com os componentes de precisão, recall e precisão. A precisão, a precisão e o recall obtidos mostraram resultados competitivos para as duas redes consideradas. Isso significa que a abordagem proposta é um classificador competitivo para os grupos de manchas solares, possibilitando a exploração e extensão de outras imagens relacionadas. No futuro, esperamos investigar mais técnicas para identificar outras características relevantes ao sol, melhorando a pesquisa nesta área espacial.

REFERÊNCIAS

CLETTE, F., BERGHMANS, D., VANLOMMEL, P., VAN DER LINDEN, R. A., KOECKELBERGH, A., & WAUTERS, L. From the Wolf number to the International Sunspot Index: 25 years of SIDC. **Advances in Space Research**, 40, 919-928. 2007.

GIOVANELLI, R. G. The Relations Between Eruptions and Sunspots. **The Astrophysical Journal**, 89, 555. 1939.

J, H. A., & KAMBER, M. Data mining: concepts and techniques. **Data mining: concepts and techniques**, (pp. 83-120). 2000.

SCHWENN, R., DAL LAGO, A., HUTTUNEN, E., & GONZALEZ, W. D. The association of coronal mass ejections with their effects near the Earth. **Annales Geophysicae**, 23, pp. 1033-1059. 2005.

SISCOE, G. The space-weather enterprise: past, present, and future. **Journal of Atmospheric and Solar-Terrestrial Physics**, 62, 1223-1232.2000.

SMOLA, A., & VISHWANATHAN, S. V. Introduction to machine learning. **Cambridge University**, UK, 32, 34. 2008.