

UTILIZAÇÃO DE TÉCNICAS DE APRENDIZAGEM DE MÁQUINA SUPERVISIONADA E NÃO SUPERVISIONADA PARA MAPEAMENTO DE ÁREA QUEIMADA

Felipe Gabriel Vieira¹; Fabiano Morelli¹; Paulo Wandernelson Pinto da Cunha¹;
Fabrício Galende Marques de Carvalho¹

¹INPE – Instituto Nacional de Pesquisas Espaciais. (felipevieiragabriel@gmail.com, fabiano.morelli@inpe.br, paulo.cunha@inpe.br, fabricio.galende@inpe.br).

Resumo: A ocorrência de queimadas representa um dos principais fatores de degradação ambiental em diferentes regiões do Brasil, especialmente em áreas de transição entre biomas, onde a diversidade da cobertura vegetal ou variações sazonais nas condições climáticas podem favorecer a ocorrência e a propagação do fogo. Nesse contexto, técnicas de aprendizagem de máquina apresentam grande potencial para automatizar processos de mapeamento de áreas afetadas por queimadas, contribuindo para o aumento da eficiência, da escalabilidade e da reprodutibilidade das análises ambientais. Assim, o presente trabalho tem como objetivo avaliar a utilização destas técnicas em modelos supervisionados e não supervisionados para o mapeamento de áreas queimadas a partir de dados de sensoriamento remoto. Na realização do estudo foram utilizadas duas imagens do satélite Sentinel-2, adquiridas com intervalo de um mês entre si. A partir das bandas B04, B08, B11 e B12 das imagens foi calculada a diferença dos índices espectrais NDVI, NBR e NBRSWIR, amplamente utilizados em análises de vegetação, com o objetivo de evidenciar alterações associadas à ocorrência de queimadas na área de estudo. Esses índices exploram diferenças na reflectância determinados intervalos, considerando que vegetações saudáveis apresentam maior reflexão na faixa do infravermelho próximo, enquanto áreas queimadas tendem a apresentar uma redução nos valores refletidos no infravermelho próximo e um aumento considerável nas bandas do vermelho e do infravermelho de ondas curtas. Posteriormente, foram utilizados dados de focos de calor registrados na região para auxiliar na definição de um ground truth empregado no treinamento do modelo de aprendizagem de máquina, contribuindo na identificação de áreas potencialmente queimadas. A partir desses critérios foi realizada a seleção manual de pontos na imagem, classificados como área queimada ou não queimada, gerando uma base de dados para o treinamento supervisionado. Em seguida, a base foi utilizada no treinamento de um modelo com o algoritmo Random Forest para realizar a classificação. Posteriormente foi aplicado um conjunto independente de dados utilizados para teste, com o objetivo de avaliar seu desempenho na identificação das áreas queimadas. Os resultados obtidos indicaram elevada capacidade de mapeamento

das áreas afetadas por queimadas, apresentando acurácia de 98,67%, precisão de 97,53% e recall de 96,20%. Além da abordagem supervisionada, também foi desenvolvido um modelo não supervisionado utilizando o algoritmo K-means para o agrupamento do conjunto de dados. Foram definidos 10 clusters com o objetivo de explorar padrões presentes nos dados, permitindo a identificação tanto de queimada e não queimadas, quanto de diferentes níveis de severidade da queima. A aplicação desse modelo possibilitou a identificação de cicatrizes de queimadas associadas aos eventos registrados na região durante o mês de observação, além de evidenciar variações internas na intensidade da queimada e mapear áreas parcialmente afetadas. De modo geral, os resultados ilustram que a integração entre índices espectrais derivados de imagens de satélite e técnicas de aprendizagem de máquina apresenta elevado potencial para o mapeamento automatizado de áreas queimadas, permitindo análises eficientes, reproduzíveis e com potencial de aplicação em diferentes contextos ambientais.

Palavras-chave: Sensoriamento remoto; Aprendizagem de máquina; Mapeamento de queimadas; Índices espectrais; Random Forest.

USING SUPERVISED AND UNSUPERVISED MACHINE LEARNING TECHNIQUES FOR BURNED AREA MAPPING

Felipe Gabriel Vieira¹; Fabiano Morelli¹; Paulo Wandernelson Pinto da Cunha¹;

Fabício Galende Marques de Carvalho¹

¹ INPE - National Institute for Spacial Reserch.. (fabiano.morelli@inpe.br , paulo.cunha@inpe.br, fabricio.galende@inpe.br).

Abstract: The occurrence of wildfires represents one of the main factors of environmental degradation in different regions of Brazil, especially in transition areas between biomes. In this context, machine learning techniques present great potential for automating the mapping of areas affected by wildfires, contributing to increased efficiency, scalability, and reproducibility of environmental analyses. Thus, the present work aims to evaluate the use of supervised and unsupervised machine learning techniques for the detection of burned areas from remote sensing data. Satellite images of the same area, acquired with an approximate interval of one month apart, were used for the study. From the spectral bands of the images, widely used spectral indices for vegetation analysis and burned area detection were calculated, namely NDVI, NBR, and NBRSWIR. These indices exploit differences in reflectance between near-infrared and shortwave infrared wavelengths, considering that healthy vegetation exhibits higher reflectance in the near-infrared range, while burned areas tend to show distinct spectral behavior due to changes in vegetation structure and soil exposure. Based on these indices, the difference between the values obtained at the two analyzed dates was calculated, with the aim of highlighting changes associated with the occurrence of wildfires in the study area. Subsequently, heat spot data recorded in the region were used to assist in defining a minimum threshold capable of indicating the presence of burned areas in each analyzed index. Based on these criteria, a manual selection of points in the image was performed, classified as burned or unburned area, generating a labeled dataset for supervised training. A model based on the Random Forest algorithm was then implemented to perform the classification, using the selected points as the basis for model training. The trained model was subsequently applied to a larger dataset in order to evaluate its performance in identifying burned areas. The results obtained indicated high detection capability for areas affected by wildfires, presenting an accuracy of 98.67%, precision of 97.53%, and recall of 96.20%. In addition to the supervised approach, an unsupervised machine learning model was also developed using the K-means algorithm for data clustering. Among these clusters, it was possible to identify classes associated with unburned areas, as well as different levels of burn severity, including regions with higher burn intensity and partially affected areas. The application of this model enabled the identification of burn scars associated with events recorded in the region during the observation month, as well as highlighting internal variations in burn intensity. Overall, the results demonstrate that the integration of spectral indices derived from satellite images and machine learning techniques presents high potential for automated burned area mapping, enabling efficient, reproducible analyses with potential application in different environmental contexts.

Keywords: Remote sensing; Machine learning; Burned area mapping; Spectral indices; Random Forest.