

AValiação DO DESEMPENHO DE MÉTODOS DE APRENDIZAGEM DE MÁQUINA PARA PREVISÃO DE INCÊNDIO EM ÁREAS DE PROTEÇÃO

Miqueias Lima Duarte¹
Jocy Ana Paixão de Sousa²
Tatiana Acácio da Silva³
Roberto Vagner Lourenço⁴

Mudanças Climáticas

Resumo

Os incêndios são eventos globais que causam perdas imensuráveis para o homem e ao meio ambiente. A previsão e o mapeamento desses eventos pode ser uma importante medida, uma vez que possibilita desenvolver estratégias com vistas ao controle ou a remediação. Desta forma, este estudo avaliou o desempenho de dois métodos de aprendizagem de máquina (*Naive Bayes* e *Deep Learning*) para previsão de incêndios em uma área localizada na região norte do Estado de Rondônia. A modelagem foi realizada a partir de um conjunto de dez variáveis explicativas, tais como a precipitação, temperatura e umidade do ar, temperatura e umidade do solo, número de dias sem chuva, distância das rodovias e o Índice de Vegetação por Diferença Normalizada. Os dados foram obtidos a partir de produtos de sensoriamento remoto processados em ambiente Sistema de Informações Geográficas. A implementação dos métodos de aprendizagem de máquina deu-se em linguagem de programação R. Os modelos foram avaliados por métricas de desempenho com base em dados independentes, o que possibilitou observar a habilidade dos métodos na predição de incêndios florestais na área de estudo. Os resultados obtidos mostraram desempenho semelhante em ambos os métodos de previsão, sendo que o *Naive Bayes* (NB) apresentou um AUC de 0,84, enquanto o *Deep Learning* (DL) apresentou um AUC de 0,82, o que representa uma pequena vantagem do NB sobre o DL. Os resultados obtidos mostraram que a integração de dados de sensoriamento remoto e a aprendizagem de máquina tem capacidade de prever o risco de ocorrência de incêndios na região.

Palavras-chave: Modelagem espacial; Aprendizagem de máquina; incêndios florestais.

¹ Aluno de Doutorado, Ciências Ambientais, Universidade Estadual Paulista (UNESP), Instituto de Ciência e Tecnologia, Sorocaba, miqueiaseng@hotmail.com.

² Aluna de Doutorado – Ciências Ambientais, Universidade Estadual Paulista (UNESP), Instituto de Ciência e Tecnologia, Sorocaba, jocy_belem@hotmail.com

³ Aluna de Mestrado – Ciências Ambientais, Universidade Estadual Paulista (UNESP), Instituto de Ciência e Tecnologia, Sorocaba, tatianacacio1909@gmail.com

⁴ Prof. Dr. Universidade Estadual Paulista (UNESP), Instituto de Ciência e Tecnologia, Sorocaba, roberto.lourenco@unesp.br.

INTRODUÇÃO

Os incêndios florestais são eventos globais resultante da associação de fatores climáticos, material combustível e atuação humana. Eles podem se originar de forma natural, bem como de forma acidental ou proposital em função da atuação antrópica. Os impactos relacionados a esses eventos afetam a saúde humana, causam prejuízos econômicos e sociais, além de alterar os ecossistemas, com perdas de áreas florestais, e depleção da qualidade do solo e água, poluição atmosférica e aumento do efeito estufa (Maeda et al., 2011).

Segundo o relatório do Painel Intergovernamental sobre Mudanças Climáticas (IPCC), o risco de incêndios florestais deve aumentar nas próximas décadas em função das mudanças climáticas e o aquecimento global. Desta forma, o delineamento de zonas de suscetibilidade a incêndios, bem como o desenvolvimento de um sistema preditivo é um passo importante, pois pode contribuir com o planejamento e gerenciamento adequado, além de possibilitar o desenvolvimento de sistemas de alerta de incêndio mais efetivo (Eastaugh e Hasenauer, 2014). Nesse contexto, os métodos de aprendizagem de máquina associado a dados de sensoriamento remoto quase em tempo real abriu oportunidades para estudos de análises quantitativas nos ecossistemas, podendo ser utilizados no monitoramento e detecção de incêndios florestais (Chuvieco et al., 2020; Pourghasemi et al., 2020).

Desse modo, esse trabalho avaliou o desempenho de dois métodos de aprendizagem de máquina associado a dados de sensoriamento remoto em uma modelagem preliminar de previsão de incêndios em uma área localizada na região norte do Estado de Rondônia.

METODOLOGIA

A área de estudo compreende a Reserva Extrativista Jaci-Paraná e a Terra Indígena Karipuna, localizado ao norte do Estado de Rondônia, com área de aproximadamente 3.510 km². Essa região vem sofrendo nas últimas décadas com o aumento de focos de incêndio devido o avanço do desmatamento pela atividade agropecuária.

O modelo de previsão de incêndios foi construído a partir de dez parâmetros

preditores, sendo eles: precipitação, temperatura e umidade do ar, temperatura e umidade do solo, número de dias sem chuva, distância das rodovias, Índice de Vegetação por Diferença Normalizada (NDVI), latitude e longitude. A precipitação (mm) foi obtida a partir de dados do *Global Precipitation Measurement - GPM*, e a temperatura de superfície (°C) e NDVI foram obtidos a partir de dados do sensor MODIS/Terra (USGS, 2020), enquanto que a umidade específica (kg.kg^{-1}), temperatura (°C) e umidade do solo (kg.m^{-2}) foram obtidos a partir de dados do Global Land Data Assimilation System-GLDAS (NASA, 2020), ambos em valores médios para o mês de agosto de 2018. A distância das rodovias foi obtida com base na vetorização de rodovias e estradas vicinais observadas em imagens do Google Earth, seguido pela geração de uma superfície contínua pelo método Distância Euclidiana. Também foram obtidos dados sobre focos de incêndio para o mês de agosto de 2018 do Banco de Dados de Queimadas do Instituto Nacional de Pesquisas Espaciais (INPE, 2020).

Os dados foram processados no software ArcGIS (ESRI, 2016), sendo adotado o sistema de coordenadas UTM, Datum SIRGAS 2000, zona 20 Sul, em seguida, os dados em *raster* foram convertidos em vetor tipo ponto para implementação dos classificadores. A previsão de incêndios florestais foi realizada a partir de classificação Bayesiana e por *Deep Learning*. A classificação Bayesiana é uma técnica de aprendizado de máquina supervisionado que estima a probabilidade de uma nova observação pertencer a uma categoria predefinida, usando um modelo probabilístico definido pela teoria de Bayes, enquanto que o método *Deep Learning* é um sistema de aprendizagem profunda (múltiplas camadas) para a extração de informação de recursos supervisionados ou não supervisionados bem como para análise de padrões (Lantz, 2019).

O processamento de dados foi realizado no software RStudio. Os dados foram divididos em conjunto de calibração (70%) e validação (30%) com auxílio do pacote *caret*. Para implementação do classificador *Naive Bayes* (NB) foi utilizado o pacote *e1071*, enquanto para a implementação do classificador *Deep Learning* (DL) foi utilizado o pacote *h2o*. A qualidade dos previsores foi avaliada por meio da acurácia, matriz de confusão, taxa de verdadeiro positivo vs falso positivo (*Receiver Operating Characteristic - ROC*) e AUC (Area Under Curve), com uso do pacote *caret* e *pROC*.

RESULTADOS E DISCUSSÃO

A Figura 1 apresenta dados sobre o desempenho de cada modelo avaliado (a), bem como a contribuição individual das variáveis para os modelos (b), e a curva ROC/AUC (c).

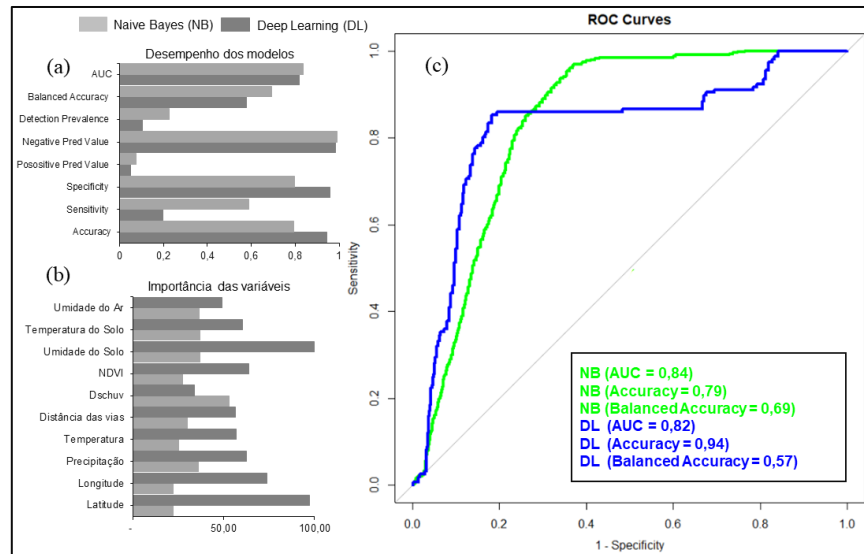


Figura 1. Estatística de desempenho individual dos modelos (a) contribuição individual das variáveis (b) e curva ROC (c).

De modo geral, observa-se que os modelos de aprendizagem de máquina utilizados apresentaram desempenho semelhantes. No que se refere-se à acurácia, o DL apresentou melhor desempenho sobre o método NB, por outro lado, quando observado os valores de AUC, o desempenho do modelo NB foi superior. O AUC representa a taxa verdadeiro-positivo contra a falso-positivo, ou seja, o número de vezes em que o classificador acertou (Fogo-Fogo) contra a o número de vezes em que o classificador errou a predição (Fogo-Não Fogo). De fato, pela matriz de confusão o número de acertos de incêndio para o NB foi de 36 para o total de 61 eventos, enquanto pelo DL foi apenas 12 acertos, essa diferença é representada pelo valor de acurácia balanceada, bem como pelo desempenho dos valores positivos previstos (Figura 1, b).

Quanto a importância das variáveis de entrada para cada modelo (Figura 1, b), observa-se que os atributos com maior contribuição para a aprendizagem profunda referem-se à umidade do solo, seguida pela latitude, longitude, NDVI e precipitação. Entretanto para o classificador *Naive Bayes* a quantidade de dias sem chuva apresentou maior importância, seguida pela precipitação, umidade do ar e do solo, e temperatura do solo.

CONSIDERAÇÕES FINAIS

Os resultados obtidos mostraram desempenhos semelhantes entre os classificadores de aprendizagem de máquina para previsão de focos de incêndio, sendo que o *Naive Bayes* apresentou uma pequena vantagem sobre o *Deep Learning*. Os resultados parciais mostraram que os métodos de aprendizagem de máquina podem ser utilizados para previsão de incêndio na região. O desempenho dos previsores pode ser aprimorado com melhoria na escala de dados utilizados, acréscimo de dados históricos sobre a ocorrência de incêndio e implementação dos algoritmos com hiperparâmetros.

AGRADECIMENTOS

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

REFERÊNCIAS

- Chuvieco, E.; Aguado, I.; Salas, J.; García, M.; Yebra, M.; Oliva, P. Satellite Remote Sensing Contributions to Wildland Fire Science and Management. *Current Forest Reports*. v. 6, p. 81-96. 2020.
- Eastaugh, C. S.; Hasenauer, H. Deriving forest fire ignition risk with biogeochemical process modelling. *Environmental Modelling & Software*. v. 55, p. 132-142. 2014.
- Instituto Nacional de Pesquisas Espaciais – INEP. Queimadas. Disponível em: <http://queimadas.dgi.inpe.br/queimadas/portal>, Acessado em: 05/07/2020.
- Lantz, B. 2019. *Machine Learning with R: Expert techniques for predictive modeling*, Third Edition. Packt Publishing. 458p.
- Maeda, E. E.; Arcoverde, G. F. B.; Pellikka, P. K. E.; Shimabukuro, Y. E. Fire risk assessment in the Brazilian Amazon using MODIS imagery and change vector analysis. *Applied Geography*. v. 31, n. 1, p. 76-84. 2011.
- NASA. 2020. GLDAS Noah Land Surface Model L4 3 hourly 0.25 x 0.25 degree V2.1, Greenbelt, Maryland, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), Acessado: 05/07/2020.
- Pourghasemi, H. R.; Geyen, A.; Lasaponara, R.; Tiefenbacher, J. P. Application of learning vector quantization and different machine learning techniques to assessing forest fire influence factors and spatial modelling. *Environmental Research*. v. 184. 2020.