

Predição de Escorregamentos de Encostas baseada em Aprendizado de Máquina

Laedson S. Pedreira ¹, Rodrigo T. Calumby ¹, Maria do S. C. São Mateus ¹

¹Universidade Estadual de Feira de Santana (UEFS)

Av. Transnordestina, s/n - Feira de Santana, Novo Horizonte - BA, 44036-900

laedsonsilva@hotmail.com, rtcalumby@uefs.br, somateus@gmail.com

Abstract. *Slope landslides are one of the main phenomena causing natural disasters in Brazilian cities, causing, every year, countless material damages and causing a large number of fatal victims. In this context, this work proposes the application of machine learning techniques to predict slope landslides, individually, in time and space, using data from multiple sources. For this, data integration was performed and several predictive algorithms about the occurrence or not of slope slides were evaluated. The results of the experiments showed that the algorithms were able to achieve promising performance.*

Resumo. *Os escorregamentos de encostas constituem um dos principais fenômenos causadores de desastres naturais nas cidades brasileiras, provocando, todos os anos, inúmeros prejuízos materiais e fazendo um grande número de vítimas fatais. Neste contexto, este trabalho propõe a aplicação de técnicas de aprendizado de máquina na predição de escorregamentos de encostas, de forma individualizada, no tempo e no espaço, a partir de dados provenientes de múltiplas fontes. Para isso, foi realizada a integração dos dados e foram avaliados vários algoritmos preditivos acerca da ocorrência ou não de escorregamentos de encostas. Os resultados dos experimentos mostraram que os algoritmos foram capazes de alcançar desempenho promissor.*

1. Introdução

Os eventos de escorregamentos de encostas ocorrem por todo planeta e vêm provocando, todos os anos, inúmeros problemas à sociedade, tais como a destruição de edificações e equipamentos urbanos, prejuízos em atividades produtivas, impactos ambientais e fazendo um grande número de vítimas fatais [IBGE 2019]. No Brasil, constituem um dos principais fenômenos causadores de desastres naturais [Tominaga et al. 2015]. Isso ocorre em função das características de seu meio físico, do seu clima tropical e da alta pluviosidade, que favorecem o desencadeamento dos escorregamentos [IBGE 2019].

Considerando a complexidade do problema e a disponibilidade de dados integrados e em larga escala, estudos específicos de modelos preditivos individualizados para monitoramento e indicação de riscos têm sido pouco explorados. Apesar disso, a aplicação de modelos preditivos baseados em aprendizado de máquina apresentam grande potencial em contribuir com ferramentas eficazes e eficientes, capazes de auxiliar no monitoramento e prevenção de danos oriundos de tais eventos.

Diante do exposto, este trabalho propõe uma modelagem preditiva no tempo e no espaço de escorregamentos de encostas, por meio da investigação e aplicação de técnicas

de aprendizado de máquina e mineração de dados, no pré-processamento e na predição, integrando os dados das ocorrências passadas com as propriedades do solo, com os dados geomorfométricos e com a precipitação pluviométrica. Além disso, busca-se avaliar múltiplos algoritmos de aprendizado de máquina.

2. Trabalhos Relacionados

Diversos estudos têm buscado utilizar as técnicas relacionadas à Inteligência Artificial para aprender a relação entre a estabilidade de encostas e suas variáveis de influência a partir de dados históricos. [Korup and Stolle 2014] analisaram 674 artigos relacionados ao tema e ressaltaram que, apesar de pesquisas aplicadas na área, a previsão de escorregamento de encostas, no tempo e no espaço, continua sendo um desafio. Os autores concluem que a maior parte dos estudos envolvem a confecção de mapas de susceptibilidade de escorregamentos, e não a previsão de riscos futuros propriamente ditos.

[Liu et al. 2021] empregaram três algoritmos (*Random Forest* (RF), *Gradient Boosted Regression Tree* (GBRT) e *MultiLayer Perceptron neural network* (MLP)) para realizar modelagem espacial de deslizamentos de terra rasos perto de Kvam, na Noruega. Foram utilizados um total de 86 deslizamentos de terra como inventário. Os autores selecionaram as amostras de não escorregamento de forma aleatória. Esse tipo de abordagem pode gerar uma configuração tendenciosa, devido a grande desproporcionalidade entre as áreas estáveis e instáveis, tornando fácil para o modelo fornecer a saída correta, e com isso maximizando as métricas de desempenho.

[Achour and Pourghasemi 2020] utilizaram três algoritmos (RF, *Support Vector Machine* (SVM) e GBRT) para prever escorregamentos de um trecho de estrada no nordeste da Argélia, com o objetivo de desenvolver mapas de suscetibilidade. Os autores obtiveram uma área sob a curva ROC (AUROC) de 0,972, e tinham como inventário apenas 28 locais de escorregamentos. Além disso, eles não levaram em consideração a chuva, que consiste no principal fator desencadeante de escorregamentos.

[Tehrani et al. 2019] utilizaram o algoritmo *Logistic Regression* (LR) para identificar escorregamentos e não escorregamentos. Os autores alcançaram uma AUROC superior a 0,88, e tinham uma base de dados com 4.542 deslizamentos de terra. Contudo, relataram que a localização dos eventos apresentavam grandes incertezas ao atribuir as coordenadas geográficas. Além disso, as amostras de não escorregamento foram criadas a partir da redução de menos de 50% das características dos casos de escorregamento. Essa abordagem pode impor um viés aos conjuntos de treinamento e teste. Além disso, utilizam um modelo geográfico de classificação da geologia de baixa precisão.

A maior parte dos trabalhos aborda uma visão estática do problema, não sendo possível prever, de fato, a ocorrência fenômeno. Além disso, os estudos não abordam amplamente questões como: predição de escorregamentos em um ambiente bem urbanizado, influenciados pela ação antrópica, além das condições ambientais; abordagem temporal na geração das amostras de não deslizamentos, visto a variabilidade das condições ambientais e a utilização das propriedades geotécnicas como fator de controle.

3. Metodologia

A base de dados usada nos experimentos foi construída a partir da integração de fontes de dados distintas. A Tabela 1 apresenta um resumo dos dados adquiridos e utilizados no

desenvolvimento desta pesquisa.

Tabela 1. Dados utilizados no experimento

heightFonte de Dado	Dado	Período
CODESAL ¹	Ocorrências de escorregamentos	2004 - 2020
Laboratório de Geotecnia da UFBA	Dados geotécnicos	1985 - 2013
SEIA ²	Precipitação pluviométrica	2004 - 2020
CEMADEN ³	Precipitação pluviométrica	2016 - 2020
INPE ⁴ (TOPODATA)	Modelo Digital de Elevação	2011

¹ Defesa Civil de Salvador; ² Sistema Estadual de Informações Ambientais e Recursos Hídricos; ³ Centro Nacional de Monitoramento e Alertas de Desastres Naturais; ⁴ Instituto Nacional de Pesquisas Espaciais.

Após a integração dos dados foi necessário realizar o pré-processamento com o intuito de remover os ruídos, adequar e codificar os dados de maneira que fosse possível usá-los no treinamento dos algoritmos. Esse processo foi realizado em etapas, sendo elas: *i*) remoção de registros incompletos e/ou duplicados; *ii*) remoção de registros de escorregamentos que não foram induzidos por chuvas; *iii*) identificação e remoção dos *outliers* através da faixa interquartil; *iv*) adequação e padronização do formato dos atributos; *v*) normalização dos dados.

Para realizar a implementação do modelo, casos que não sejam de escorregamentos devem ser adicionados ao conjunto de amostras, tendo assim, duas classes possíveis: positiva (escorregamento) e negativa (não escorregamento). Para isso, gerou-se um registro de não ocorrência para cada registro de escorregamento, sendo que cada registro de não escorregamento se deu 96 horas anterior à ocorrência do evento de deslizamento.

Ao final dessas etapas, o conjunto de dados consistiu em 7.224 registros pertencentes a duas classes: *i*) escorregamento: 3.612 registros; *ii*) não escorregamento: 3.612 registros. Para o treinamento/validação foi separada 80% da base de dados, restando 20% para os testes. O treinamento foi realizado utilizando a técnica *10-folds-cross-validation*, com dez repetições.

Foram implementados os seguintes algoritmos de aprendizado de máquina: *Decision Trees* (DT), LR, RF, *Naive Bayes* (NB) e MLP. A Tabela 2 apresenta os hiperparâmetros utilizados em cada um dos classificadores. Como se trata de um experimento inicial, a escolha desses parâmetros foi por meio de tentativas exploratórias. A avaliação da eficácia foi realizada a partir das medidas de acurácia (Acc) e AUROC. Os modelos foram escolhidos com base na média das AUC das amostras de validação.

Tabela 2. Parâmetros utilizados nos classificadores escolhidos

Classificador	Parâmetros
Decision Trees	criterion = entropy, max_depth = 100
Logistic Regression	max_iter = 1000, C = 8
Random Forest	n_estimators=7000, criterion=gini, min_samples_split=2
Naive Bayes	Nenhum
Multilayer Perceptron	hidden_layer_sizes = 300, max_iter = 20000

4. Resultados

As métricas de avaliação Acc e AUROC foram calculadas para as fases de treinamento e teste. Como pode ser analisado na Tabela 3, o algoritmo RF apresentou o melhor desem-

penho entre os cinco algoritmos analisados. Além disso, os resultados computados com as medidas de avaliação se mantiveram estáveis entre as etapas de treinamento e teste, denotando uma boa capacidade de generalização do modelo.

Tabela 3. Resultados obtidos nos experimentos

Algoritmo	Treinamento		Teste	
	Acc	AUROC	Acc	AUROC
Decision Trees	0,7362	0,7364	0,7347	0,7347
Naive Bayes	0,6760	0,6760	0,6690	0,6689
Random Forest	0,8028	0,8029	0,8081	0,8066
Logistic Regression	0,6746	0,6748	0,6835	0,6837
Multilayer Perceptron	0,7133	0,7143	0,7258	0,7220

5. Considerações Finais

Este trabalho abordou o uso de algoritmos de aprendizado de máquina para predição de escorregamentos de encostas. Os resultados apresentaram uma acurácia superior 80% e AUROC superior 0,80, nas fases de treinamento e teste. Embora a estrutura de previsão que foram relatados neste estudo estejam na sua fase inicial, os resultados já demonstram que os modelos podem ser utilizados para prever deslizamentos de terra.

Como trabalhos futuros pretende-se utilizar o *gridsearch* com o objetivo de obter os melhores hiperparâmetros, otimizando o desempenho dos classificadores. Além disso, pretende-se explorar outros algoritmos de aprendizado de máquina. Por fim, serão estabelecidos cenários de análise, com o intuito de verificar o impacto de cada um, em termos de poder preditivo. Esses cenários serão desenvolvidos com simulações a partir dos dados, tais como: i) predição variando dados acerca da previsão do tempo; ii) predição variando dados acerca das propriedades do solo; iii) criação de amostras de não escorregamentos através da localização de casos reais onde não ocorreu a instabilidade da encosta.

Referências

- Achour, Y. and Pourghasemi, H. R. (2020). How do machine learning techniques help in increasing accuracy of landslide susceptibility maps? *Geoscience Frontiers*, 11(3):871–883.
- IBGE (2019). *Suscetibilidade a deslizamentos do Brasil : primeira aproximação*. Instituto Brasileiro de Geografia e Estatística - IBGE, Rio de Janeiro.
- Korup, O. and Stolle, A. (2014). Landslide prediction from machine learning. *Geology Today*, 30.
- Liu, Z., Gilbert, G., Cepeda, J. M., Lysdahl, A. O. K., Piciullo, L., Hefre, H., and Lacasse, S. (2021). Modelling of shallow landslides with machine learning algorithms. *Geoscience Frontiers*, 12(1):385–393.
- Tehrani, F. S., Santinelli, G., and Herrera, M. (2019). A framework for predicting rainfall-induced landslides using machine learning methods. *17th European Conference on Soil Mechanics and Geotechnical Engineering, ECSMGE 2019 - Proceedings*.
- Tominaga, K., Santoro, J., and do. Amaral, R. (2015). *Desastres naturais: conhecer para prevenir*, volume 3ed. Instituto Geológico, São Paulo.