
ÁRVORE DE DECISÃO APLICADA NA IDENTIFICAÇÃO DE ÁREAS DE RISCO DE DESLIZAMENTO: COMPARAÇÃO DOS MÉTODOS J48, *FUNCTIONAL TREES* E *RANDOM TREE*

ALINE BERNARDA DEBASTIANI (1)
MAURÍCIO DE SOUZA (1)
MARCIELLI AP. BORGES DOS SANTOS (1)
RICARDO DAL'AGNOL DA SILVA (2)
GUILHERME MOACIR SCHWADE (1)
LUÍSA LELIS (1)
MOSAR FARIA BOTELHO (3)

(1) Universidade Tecnológica Federal do Paraná
Acadêmicos do curso de Engenharia Florestal, Dois Vizinhos – PR
{aline_ck; dark_mds; marcielle_bs; guilherme_schwade}@hotmail.com; lelis.luisa@yahoo.com.br

(2) Instituto Nacional de Pesquisas Espaciais – INPE
Mestrando em Sensoriamento Remoto
silvard@dsr.inpe.br

(3) Universidade Tecnológica Federal do Paraná
Prof. Dr. Mosar Faria Botelho, Dois Vizinhos - PR
mosar@utfpr.edu.br

Fenômenos naturais como tornados, terremotos, tsunamis e deslizamentos de encosta, ocorrem no mundo inteiro e de forma cada vez mais frequente. Diante disso, o desenvolvimento de metodologias adequadas e o uso de tecnologias, surgem como uma ferramenta para auxiliar na prevenção e na tentativa de evitar a ocorrência de tais eventos. As árvores de decisão são estruturas capazes de aprender e tomar decisões a partir de um conjunto de amostras conhecidas, onde a partir deste aprendizado outras amostras são classificadas, tendo a vantagem de possuírem fácil compreensão (POZZER, 2006). O software WEKA possui vários algoritmos de árvore de decisão, entre eles o J48, *Functional Trees* (FT) e *Random Tree*. O J48 é a versão open source do algoritmo de árvore de decisão C4.5, o qual é um classificador estatístico baseado no conceito de entropia, onde a cada nó, o algoritmo determina qual atributo realiza a melhor separação entre classes, gerando uma árvore univariada (onde cada regra é baseada em somente um atributo). O FT é um algoritmo de construção de árvores de decisão oblíquas, onde as árvores são multivariadas – cada regra possui uma expressão baseada em mais de um atributo. O algoritmo *Random Tree* considera apenas alguns atributos escolhidos aleatoriamente para cada nó da árvore, não realiza poda e também não permite a estimativa de probabilidade de classes. O objetivo do presente estudo é de comparar os algoritmos de árvore de decisão J48, FT e *Random Tree* para identificação de áreas suscetíveis a deslizamentos. Como atributos para os classificadores, foram utilizadas as informações de uso e cobertura da terra, imagem orbital, classe de solo, declividade, direção do fluxo, comprimento do fluxo, hidrografia e mapeamento das ocorrências de deslizamento. A área de estudo desse trabalho foi a região litorânea do estado do Paraná, onde foi utilizado de dados de referência de campo para a ocorrência de deslizamentos obtidos pela Mineropar em 2011. Esses dados foram utilizados para geração dos modelos e validação dos resultados. Para gerar o mapa de uso e cobertura da terra foi utilizada a imagem ortorretificada do mosaico Landsat Geocover e aplicada uma classificação supervisionada por Máxima verossimilhança. A partir do modelo digital de elevação ASTER GDEM foi derivada a informação de declividade, e a partir desta foi gerada a direção do fluxo de água e comprimento de fluxo do escoamento da água. A informação da classe de solo foi extraída do primeiro nível categórico do levantamento realizado pela Embrapa em 2006. Também foi utilizado o arquivo de reservatórios de água, denominado como Rios Massa d'água, gerado pela Copel e Águas Paraná em 2011. Todos os dados foram projetados em UTM, Datum WGS 1984, Zona 22 sul, reamostrados para pixels de 5 m e recortados de forma a abranger a área de estudo. Buscou-se o mapeamento temático das classes de: ocorrência e não ocorrência de deslizamento. Portanto, foram coletadas 55.452

amostras de ocorrência e não ocorrência a partir de amostragem aleatória, estes dados foram destinados para o treinamento da árvore de decisão. O treinamento foi realizado utilizando 30% das amostras e a validação foi realizada utilizando os demais 70%. A avaliação da qualidade da predição para os métodos levou em consideração o índice kappa e a exatidão global. O índice kappa varia de -1 a 1, onde o valor 1 indica completa concordância entre a referência e a predição, 0 indica completa aleatoriedade, e -1 indica completa discordância. A exatidão global expressa o percentual de acerto da predição com relação a referência. Testou-se a concordância do índice kappa das classificações através do teste estatístico Z proposto por Congalton e Mead (1983), quando o índice kappa é significativo de acordo com o teste Z, indica que há concordância na classificação. O teste estatístico Z ainda foi utilizado para comparar se há diferenças significativas entre os classificadores. Os algoritmos FT e J48 obtiveram índice kappa de 0,76 e exatidão global de 88,31%, enquanto o algoritmo *Random Tree* obteve um resultado inferior, com kappa de 0,75 e 86,74% de exatidão global. O índice kappa obtido nas classificações foi significativo, sendo ainda que o kappa dos algoritmos FT e J48 foram significativamente maior que do algoritmo *Random Tree* de acordo com o teste estatístico Z. Apesar do resultado dos algoritmos FT e J48 terem sido iguais, o algoritmo J48 possui certo destaque por ser mais intuitiva, pois proporciona um maior entendimento de como acontece a separação de classes. Os métodos de árvore de decisão avaliados mostraram-se ferramentas com alto potencial para identificação de áreas de risco de deslizamento, além de serem métodos que apresentam custo computacional baixo comparado à outras ferramentas de inteligência artificial, com reprodutibilidade garantida e estarem disponíveis em software livre.

Agradecimentos

Agradecimentos à Fundação Araucária que tem possibilitado a realização desse projeto através da concessão das bolsas de iniciação científica além do apoio para sua divulgação no Colóquio Brasileiro de Ciências Geodésicas.

Referências

Congalton, R. G.; Mead R. A. A quantitative method to test for consistency and correctness in photointerpretation. **Photogrammetric Engineering and Remote Sensing**, v.49, n.1, p.69-74, 1983.

POZZER, Cesar Tadeu. **Aprendizado por Árvores de Decisão**. Universidade Federal de Santa Maria, Rio Grande do Sul, 2006.