

AJUSTE DE HIPERPARÂMETROS DE UMA REDE NEURAL MEDIANTE ALGORITMO GENÉTICO NA GERAÇÃO DE FUNÇÕES DE PEDOTRANSFERÊNCIA QUE ESTIMAM A RETENÇÃO DE ÁGUA NO SOLO

THIAGO REIS PORTO¹; MIGUEL DAVID FUENTES-GUEVARA²; ULISSES BRISOLARA CORRÊA³; LUÍS CARLOS TIMM⁴

¹Universidade Federal de Pelotas – tporto@inf.ufpel.com.br

²Universidade Federal de Pelotas – miguelfuge@hotmail.com

³Universidade Federal de Pelotas – ub.correa@inf.ufpel.edu.br

⁴Universidade Federal de Pelotas - lctimm@ufpel.edu.br

1. INTRODUÇÃO

Redes Neurais Artificiais (RNAs) podem ser empregadas para desenvolver funções de pedotransferência (FPTs) que permitem obter estimativas de diversos atributos do solo (HAGHVERDI, A.; CORNELIS, W. M.; GHAMRAN, B., 2012). Neste trabalho, uma RNA é utilizada para gerar FPTs em solos da bacia hidrográfica do Arroio Pelotas, mais precisamente, objetivando estimar a retenção de água no solo. O interesse na geração de FPTs é devido a que a análise de atributos físico-hídricos do solo, como a retenção de água no solo por métodos tradicionais (e.g. câmara de Richards) requer árduo trabalho em campo e/ou laboratório, tornando a determinação deste atributo demorada e onerosa. FPTs conseguem prever atributos do solo de difícil determinação através de dados de solo obtidos corriqueiramente no laboratório. Porém, muitas FPTs desenvolvidas na atualidade foram criadas para prever a retenção de água em solos de condições de clima diferente dos encontrados no Brasil, dificultando o uso de FPTs já desenvolvidas devido às diferenças na composição de solos de diferentes climas (TAGAMI, TWINING & WASSERMAN, 2012).

Em Aprendizado de Máquina (AM), existe a necessidade de uma política de redução de dimensionalidade e busca de hiperparâmetros, uma vez que um número grande de atributos costuma deixar os algoritmos lentos ou menos precisos, e é um dos motivos que redes neurais profundas (que lidam melhor com esse problema) são preferíveis a redes rasas (POGGIO, T., MHASKAR, H., ROSASCO, 2017). Hiperparâmetros são parâmetros do próprio algoritmo, que definem certas propriedades de como o algoritmo deve treinar o modelo. Um excesso de atributos e hiperparâmetros, também torna a complexidade de tempo na exploração do espaço de busca exponencial, portanto o uso de estratégias de heurística ou meta-heurística para não realizar uma busca exaustiva e demorada é interessante. Uma opção plausível são os Algoritmos Genéticos (AGs), algoritmos probabilísticos e baseados em populações que funcionam como meta-heurística, inspirados na Teoria da Evolução realizam uma busca paralela e adaptativa em um espaço de soluções (PACHECO et al., 1999). Frente ao anteriormente exposto, o objetivo deste trabalho foi avaliar o potencial do AG em ajustar os hiperparâmetros de uma RNA desenvolvida para gerar FPTs que estimem a retenção de água nos solos de uma bacia hidrográfica.

2. METODOLOGIA

Este trabalho utilizou um banco de dados de solo obtido da coleta e análise de amostras de solo na Bacia Hidrográfica do Arroio Pelotas (BHAP) e de uma de

suas sub-bacias a Sub-Bacia Hidrográfica Sanga Ellert (BHSE). A BHAP é de extrema importância para o abastecimento de água e produção agrícola da região, em uma área de 910km² se estende por quatro municípios: Canguçu, Morro Redondo, Arroio do Padre e Pelotas (MEGIATO;MOURA, 2019). Amostras foram coletadas ao longo de um transecto espacial(neste caso um caminho linear ao longo da área) na bacia devido a sua grande extensão, enquanto na sub-bacia os dados foram coletados em malha uniformemente cobrindo a maioria da região da sub-bacia por possuir menor área (cerca de 0,7 km²).

A RNA avaliada possui três camadas (entrada, oculta e saída), escrita em Python com uso da biblioteca PyTorch, é uma rede para prever a retenção de água no solo na BHAP. Os dados de entrada possuem 10 atributos do solo sendo eles: percentagem de argila, silte e areia, densidade, macroporosidade, microporosidade, carbono orgânico, potencial matricial aplicado (h) e a água retida no respectivo h . Os dados também contém 6 atributos topográficos: elevação do terreno, declividade, aspecto do terreno, curvatura do terreno, curvatura do plano e curvatura do perfil. A rede aceita entradas com diversos valores de potencial matricial, e gera uma saída aproximada com um valor contínuo do valor de umidade. Esse modelo de rede para desenvolver FPTs é baseado nos trabalhos de Haghverdi, Cornelis e Ghahraman (2012) e Haghverdi, Öztürk e Cornelis (2014) que utilizaram um método que denominam Rede Neural Pseudo-contínua para gerar FPTs.

Para o ajuste dos hiperparâmetros e a redução da dimensionalidade foi utilizado inicialmente a busca Random Search (RS - BERGSTRÄ, James; BENGIO, Yoshua, 2012). Essa técnica, embora explore melhor o espaço do que uma busca exaustiva, ainda pode ser bastante demorada para se encontrar algum resultado satisfatório. Portanto, com o interesse de explorar alternativas que reduzissem o tempo de busca, e melhorasse os resultados da rede, foi escolhido o Algoritmo Genético (AG). Para aplicarmos o AG na busca, foram utilizadas as bibliotecas de Python: sklearn genetic-opt, DEAP (*Distributed Evolutionary Algorithms in Python*), scikit-learn e PyTorch (para a RNA). Para aplicação do AG, a entrada com os atributos do solo e hiperparâmetros, é codificada em binário, para possibilitar operações com operadores binários, no cruzamento e mutação do AG. No cruzamento, dois indivíduos pais (nesse caso modelos da RNA) são escolhidos, e baseado em cortes aleatórios são divididos e recombinados, formando novos indivíduos. Isso é semelhante ao cruzamento cromossômico biológico, por isso os indivíduos também são chamados cromossomos no contexto de AG. Foi aplicado neste projeto uma probabilidade de cruzamento na população de uma geração de 80%. O método utilizado para a seleção dos pais foi o torneio com tamanho 3, onde 3 indivíduos são escolhidos baseados em uma probabilidade (com base na avaliação) e competem entre si, a melhor avaliação vence (é selecionada). Foram testadas diversas configurações neste modelo de AG, e então foram coletados alguns dados e gráficos para comparações com resultados anteriores, que serão discutidos na próxima sessão.

3. RESULTADOS E DISCUSSÃO

Os resultados obtidos até o momento mostraram o AG como uma alternativa viável para busca de hiperparâmetros e redução de dimensionalidade. Destacando que o AG maximizou a métrica de precisão R^2 durante as gerações de população, e reduziu os valores das métricas de erro absoluto médio (MAE - do Inglês Mean

Square Error) e erro quadrado médio ao quadrado (RMSE - do inglês Root Mean Squared Error). O AG obteve resultados com uma melhora interessante na generalização da rede, quando comparado com os resultados obtidos *a priori* da busca com RS. O AG também pode ser executado em menos tempo, desde que não seja utilizado um número muito grande de população e/ou gerações. Nos primeiros testes, o AG foi utilizado para um ajuste do modelo resultado do RS, realizando uma busca apenas nos atributos de entrada (redução de dimensionalidade) e com os mesmos hiperparâmetros. Nesta comparação, o RS reduziu os atributos para argila, densidade, microporosidade, carbono, curvatura do terreno e curvatura do plano, e o AG reduziu os atributos em argila, densidade, macroporosidade, declividade e curvatura do plano. Estes primeiros testes já demonstraram que possivelmente as características do solo como teores de argila, assim como a porosidade do solo e a alguns atributos topográficos são importantes para descrever a capacidade de retenção de água no solo com a RNA.

A Figura 1 contém os resultados de um dos melhores modelos encontrados com o AG, executado com população de cerca de 50 indivíduos e 100 gerações, e também passou por um pequeno tratamento nos dados (normalização) que ajudou a melhorar a avaliação da rede. Este modelo, porém utiliza mais atributos que os testes anteriores (argila, silte, areia, microporosidade, carbono orgânico do solo, elevação, aspecto do terreno, curvatura do plano e curvatura do perfil). Também com o *scatter plot* (Figura 1 a direita) se nota que embora as previsões estejam bem distribuídas em torno da diagonal dos valores esperados, a variância ainda está bastante alta, especialmente para os valores mais altos de h , isso também aconteceu nos resultados anteriores da RNA.

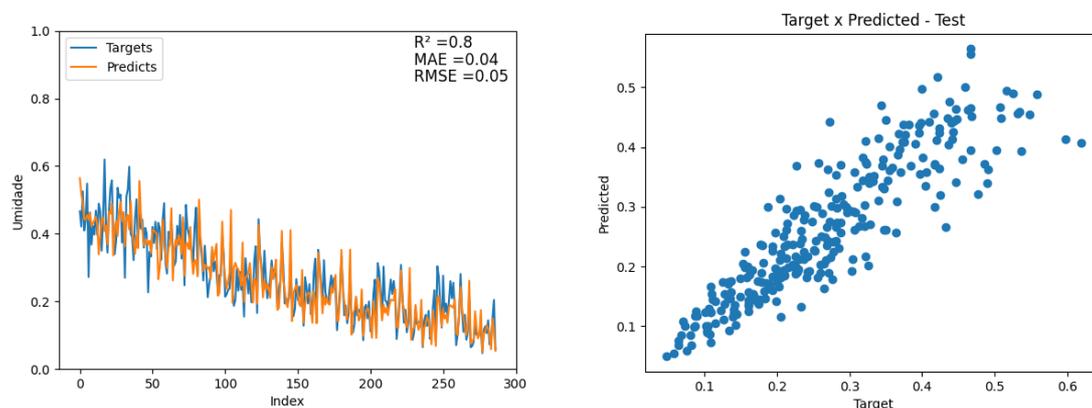


Figura 1. Resultados de modelos encontrados com AG no conjunto teste.
Esquerda: Plot - retenção de água no solo x Índice (Indica a ordem no plot).
Direita: Scatter plot - Estimado x Valor esperado.

4. CONCLUSÕES

O AG se mostrou promissor como alternativa ao Random Search. Os resultados ajudaram a identificar os atributos do solo e atributos topográficos possivelmente mais relevantes da base de dados, para predição da retenção de água no solo. A maioria dos melhores modelos utilizaram algum atributo físico do solo como, porosidade do solo, teor de argila, areia e silte, e algum atributo topográfico (e.g. aspecto, curvatura e elevação). Os resultados demonstraram alta variabilidade nas previsões de forma acentuada nas amostras com maiores

potenciais matriciais, indicando que ainda precisa ser investigado. Esse fenômeno potencialmente pode ser uma deficiência dos modelos testados, ou devido às características dos dados de entrada.

5. REFERÊNCIAS BIBLIOGRÁFICAS

BARROS, A. H. C.; LIER, Q. d. J. van. **Pedotransfer functions for Brazilian soils. In: Application of Soil Physics in Environmental Analyses.** [S.l.]: Springer, 2014. p.131–162

BERGSTRA, James; BENGIO, Yoshua. **Random search for hyper-parameter optimization. Journal of machine learning research,** v. 13, n. 2, 2012.

FUENTES-GUEVARA, M. D. **Modelagem da retenção de água em solos sob clima subtropical usando funções de pedotransferência e redes neurais artificiais.** Tese de Doutorado (Doutor em Ciências – PPG MACSA), 2022. 158p.

HAGHVERDI, A.; CORNELIS, W. M.; GHAHRAMAN, B. **A pseudo-continuous neural network approach for developing water retention pedotransfer functions with limited data.** *Journal of Hydrology*, v. 442–443, p. 46–54, 2012. Disponível em: <<http://dx.doi.org/10.1016/j.jhydrol.2012.03.036>>

HAGHVERDI, A.; ÖZTÜRK, H. S.; CORNELIS, W. M. **Revisiting the pseudo continuous pedotransfer function concept: Impact of data quality and data mining method.***Geoderma*, v. 226–227, n. 1, p. 31–38, 2014. Disponível em:<<http://dx.doi.org/10.1016/j.geoderma.2014.02.026>>

MEGIATO, É. I.; MOURA, N. S. V. **Mapeamento geomorfológico da Bacia Hidrográfica do Arroio Pelotas, RS.** *Geografia (Londrina)*, [S.l.], v.28, n.1, p.151–163, 2019.

PACHECO, M. A. C. et al. **Algoritmos genéticos: princípios e aplicações. ICA: Laboratório de Inteligência Computacional Aplicada.** Departamento de Engenharia Elétrica. Pontifícia Universidade Católica do Rio de Janeiro. Fonte desconhecida, [S.l.], v.28, 1999.

POGGIO, T., MHASKAR, H., ROSASCO, L. et al. **Why and when can deep-but not shallow-networks avoid the curse of dimensionality: A review.** *Int. J. Autom. Comput.* **14**, 503–519 (2017). Disponível em: <<https://doi.org/10.1007/s11633-017-1054-2>>

TAGAMI, Keiko; TWINING, John R.; WASSERMAN, M. Angelica V. **Terrestrial radioecology in tropical systems. In: Radioactivity in the Environment.** Elsevier, 2012. p. 155-230.