



II CONFERÊNCIA

DE GERENCIAMENTO DE ÁREAS CONTAMINADAS

30 de agosto a 01 de setembro de 2022

Sessão: Big Data.

Potencial Aplicação de Redes Neurais Artificiais (ANN) na Predição da Eficiência de Processos de Remediação Ambiental

Murilo Gonçalves da Rocha^{1*}

¹LabGEO Laboratório e Pesquisa, Fazenda Rio Grande, Brasil

*Autor correspondente: murilo@labgeo.com.br

Palavras-chave: big data, redes neurais artificiais, predição.

Resumo:

Com a ascensão da era digital, a humanidade passou a produzir uma quantidade extremamente grande de dados. Essa grande quantidade de dados guarda muitas informações e correlações potencialmente úteis que podem ser extraídas mediante o tratamento estatístico adequado, mas o grande volume de dados a serem tratados pode dificultar esse processo. Uma das maneiras mais promissoras de extrair *insights* valiosos de dados em grande escala é a implementação de algoritmos de inteligência artificial. Em particular, métodos de *deep learning* são bastante úteis, pois seu poder preditivo aumenta conforme a quantidade de dados disponíveis também aumenta, ao passo que algoritmos tradicionais de aprendizado de máquina tendem a estagnar em poder preditivo. Um exemplo de algoritmo de *deep learning* são as redes neurais artificiais (ANN - *artificial neural network*), uma forma de inteligência artificial baseada em neurônios artificiais, que se comportam como análogos eletrônicos de um cérebro biológico. No contexto da remediação ambiental, uma possível aplicação da inteligência artificial é na predição da eficácia de possíveis projetos de remediação ambiental, baseando-se em dados coletados por empresas do ramo ao longo de um histórico de intervenções realizadas. O objetivo do presente trabalho é demonstrar como a implementação de redes neurais artificiais para a análise de dados pode ser útil nesse contexto.

Para exemplificar esse processo, foi construído um banco de dados artificial com 1000 amostras, cada uma representando um projeto de intervenção ambiental onde foi feita a oxidação química *in situ* do tetracloroetileno, utilizando persulfato de sódio com diferentes ativadores. As variáveis que impactam na atenuação da concentração de tetracloroetileno foram a concentração de oxidante injetada, o pH da água do aquífero contaminado, as concentrações de carbonato e cloreto dissolvidos no aquífero, o tipo de solo (arenoso, argiloso, areno-argiloso e siltoso), o tipo de ativador utilizado em conjunto com o oxidante (ferro quelado, base ou calor) e a concentração do contaminante no aquífero. Além disso, há também a interação entre variáveis (como, por exemplo, uma queda significativa na eficiência quando há muito cloreto dissolvido e a forma de ativação é ferro quelado). Foi tomado o cuidado de produzir um dataset cujas influências das variáveis são quimicamente plausíveis, além do fato de se introduzir ruído aleatório nas respostas, para eliminar o determinismo absoluto de como as porcentagens de degradação dependem das variáveis, aproximando o *dataset* de uma situação real. Esse *dataset* foi então utilizado como *input* para treinar uma rede neural artificial criada na linguagem de programação Python, com o intuito de prever como a porcentagem de tetracloroetileno degradada depende de cada uma das variáveis listadas.

A figura 1 mostra como o erro quadrático médio (MSE) diminuiu conforme a rede neural iterava sobre o *dataset*, aprendendo os padrões e correlações entre a variável dependente e as independentes. É possível observar que antes de 10 iterações, o modelo já havia convergido para um erro quadrático médio de cerca de 1%, tanto nos dados de treino quanto nos dados de teste, demonstrando que a rede neural aprendeu as relações entre as variáveis sem apresentar problemas de *overfitting* (aprender erros aleatórios como se fossem características do fenômeno).

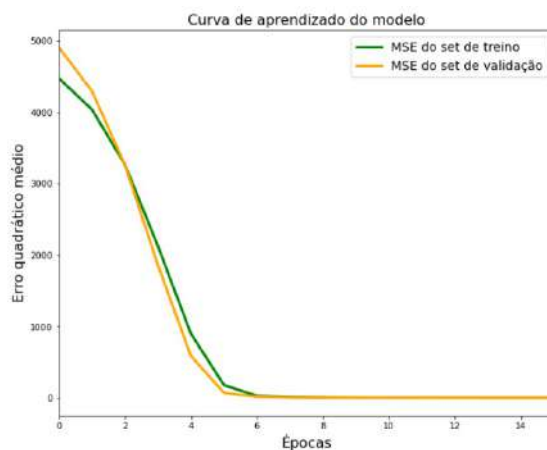


Figura 1: curva de aprendizado do modelo, mostrando que menos de 10 iterações foram o suficiente para a rede neural aprender como a eficiência da degradação de tetracloroetileno depende das variáveis listadas.

Na figura 2 estão plotados os valores preditos pelo modelo contra os valores reais. Para um modelo 100% preciso, esse gráfico mostraria uma linha reta. O fato de esse gráfico se aproximar de uma reta mostra que os valores preditos estão muito próximos do real. Para esse modelo em particular, o erro médio absoluto das predições é de $\pm 1,0644\%$ de degradação do tetracloroetileno.

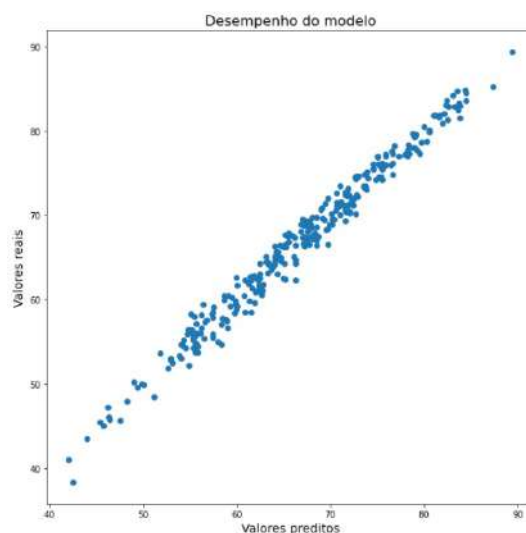


Figura 2: valores reais plotados contra os valores preditos, mostrando um bom poder preditivo.

De posse de uma rede neural treinada, é possível fazer predições acerca da eficácia da remediação com base nas informações coletadas no *site* a ser tratado. Por exemplo, tendo os valores de concentração de cloreto e carbonato no aquífero, bem como o tipo de solo e a concentração de tetracloroetileno presente no local, é possível prever qual a porcentagem do contaminante seria degradada caso escolhêssemos injetar persulfato de sódio a diferentes concentrações e ativado de diferentes maneiras. Isso permitiria uma escolha melhor de quais são as tecnologias mais apropriadas para a intervenção ambiental em questão, tanto em termos de atenuação da concentração dos contaminantes quanto em termos financeiros.

Nesse sentido, o modelo permite que várias abordagens sejam exploradas antes da intervenção propriamente dita, possibilitando uma escolha mais rápida e assertiva de como tratar o local em questão, otimizando tanto a eficácia no tratamento quanto a viabilidade econômica do projeto.

Referências

Gasteiger, J. (1993). Neural Networks in Chemistry. *Angewandte Chemie*, 32, 4, 503-527.

Liang, C. (2006). Influences of carbonate and chloride ions on persulfate oxidation of trichloroethylene at 20 °C. *Science of The Total Environment*, 370, 271-277.

Bennedsen, L. Influence of chloride and carbonates on the reactivity of activated persulfate. *Chemosphere*, 86, 1092-1097.