NOVO MÉTODO DE ANÁLISE DE DADOS DE POUCAS AMOSTRAS DE SOLO BASEADO EM DEEP LEARNING

Gabriel Tonon Cimatti – Universidade Estadual de Ponta Grossa – UEPG – 22001926@uepg.br
Alaine Margarete Guimarães – Universidade Estadual de Ponta Grossa – UEPG – alainemg@uepg.br
Gabriel Passos de Jesus – Universidade Estadual de Ponta Grossa – UEPG – 3100122004015@uepg.br

Resumo: Uma parte essencial para produtividade na agricultura de precisão é a saúde do solo. O carbono orgânico total é um grande indicador dessa saúde e desenvolver um método rápido e acessível para estimar esse indicador é de extrema importância. Esse trabalho tem como objetivo desenvolver um método para a predição do carbono orgânico total no solo utilizando dados obtidos através da técnica de espectroscopia e métodos de machine learning, em específico, deep learning.

Palavras-chave: Agricultura de precisão, Deep Learning, aprendizado de máquina, análise de dados agrícolas.

1. Introdução

Na agricultura de precisão, o carbono orgânico total (COT) do solo é um grande indicador da saúde daquele solo porém chegar nesse valor atualmente requer métodos caros e demorados. Uma solução para esse problema é análise de amostras do solo por espectroscopia, que consiste em correlacionar a resposta espectral de amostras do solo com seus componentes físico-químicos. O desafio dessa abordagem é a pequena quantia de amostras e a alta dimensionalidade do problema. Por esses fatores citados esse trabalho tem como objetivo a predição do COT no solo com base nos índices de refletância obtidos através de amostras do solo expostas a diferentes faixas de radiação infravermelho.

2. Materiais e Métodos

Os dados para essa pesquisa foram obtidos a partir de 252 amostras de solo, coletadas em diferentes profundidades, na Fundação ABC, no município Ponta Grossa, PR.

Dessas amostras foram medidos os índices de refletância após serem expostas a diferentes comprimentos de onda de raios infravermelho. Na base foi aplicados 5 algoritmos de seleção de atributos, com o intuito de reduzir a dimensionalidade do problema. Após isso foram geradas diversas bases com subconjuntos dos dados originais, estas foram alimentadas a algoritmos de deep learning e a métodos estatísticos convencionais. Os algoritmos escolhidos foram a regressão de mínimos quadrados parciais, a rede neural convolucional e a rede neural de memória de curto longo prazo. Os algoritmos para essas tarefas foram escolhidas com base nas pesquisas de Xueying li, et al. e Zijin Bai, et al.

3. Resultados

O método estatístico Regressão por Mínimos Quadrados Parciais se mostrou a pior solução para o problema, com a raiz quadrada do erro médio variando entre 33 até 35. Ao utilizar algoritmos de deep learning foi notável a diferença favorável. Utilizando o algoritmo de memória de curto longo prazo a métrica variou entre 9.2 e 9.3 e com o algoritmo de rede neural convolucional a métrica ficou em 8.9.

4. Conclusão/ões

Foi concluido que os algoritmos de deep learning se mostraram superiores ao método convencional estatístico na predição do COT e que a redução da dimensionalidade do problema através de algoritmo ajudou a melhorar a acurácia da rede final.

Agradecimentos

Esta pesquisa foi apoiada pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico.

Referências

Li, Xueying et al. Optimizing soil carbon content prediction performance by multi-band feature fusion based on visible near-infrared spectroscopy. **Journal of Soils and Sediments**. 2024.

Bai, Zijin et al. Estimation of soil inorganic carbon with visible near-infrared spectroscopy coupling of variable selection and deep learning in arid region of China. **Geoderma**. 2023.