ASSOCIAÇÃO DE DADOS ESPECTRAIS E TÉCNICAS DE APRENDIZADO DE MÁQUINA PARA ESTIMATIVA DO TEOR DE CARBONO ORGÂNICO TOTAL NO SOLO

Caroline Heloíse de Oliveira (BIC/FA), Gabriel Passos de Jesus, Alaine Margarete Guimarães (Orientadora), e-mail: alainemg@uepg.br.

Universidade Estadual de Ponta Grossa/Departamento de Informática.

Ciências Exatas, Ciência da Computação

Palavras-Chave: *Mineração de dados, Espectroscopia, Solos Tropicais, Alta dimensionalidade, Seleção de atributos.*

Resumo

Os métodos tradicionais para obtenção de teor de carbono apresentam diversos pontos negativos. A espectroscopia é uma alternativa eficiente, porém necessita de métodos computacionais para a realização das análises. Neste trabalho foram realizados testes para determinar os melhores algoritmos para essa determinação. Os resultados foram favoráveis e apontam as melhores configurações para esse caso.

Introdução

O teor de Carbono Orgânico Total (TOC) é uma temática importante para o meio ambiente e para a agricultura (Angelopoulou *et al*.,2020). Os métodos tradicionais para determinação de TOC são onerosos, demandam tempo e precisam de reagentes químicos altamente poluentes. A espectroscopia é uma alternativa eficaz, rápida, de baixo custo e limpa (Barra, 2021). Contudo, este método necessita de análises computacionais. Técnicas de Aprendizado de Máquina (AM), incluindo a seleção de atributos, são adequadas para esta tarefa, podendo estabelecer modelos eficientes de estimativa de teor de atributos no solo.

O objetivo desse trabalho foi aplicar técnicas de AM usando algoritmos de seleção de atributos e de regressão, para estimar o teor de TOC em amostras de solo da região dos Campos Gerais – Paraná.

Materiais e Métodos

Foram analisadas 252 amostras de solo de um experimento de sistema de rotação de culturas de longa duração. As amostras foram coletadas em 28 pontos diferentes e em 9 profundidades. O TOC foi determinado pelo método de Pavan *et al* (1992). As análises foram realizadas no equipamento Espectrofotômetro FOSS NIRS DA1650 que abrange a faixa de 1100 nm a 1650 nm na região do infravermelho próximo, realizando os registros a cada 2 nm em absorbância.

Em seguida, foram criados, na linguagem de programação Python, *scripts* para o tratamento dos dados obtidos pelo equipamento, convertendo valores de absorbância para refletância e transmitância e organizando os dados em um arquivo do tipo CSV junto aos valores de TOC. O *script* também apresenta a possibilidade de transformações da base de dados.

Para a construção do modelo, foi utilizado o *framework* WEKA versão 3.8.6. Foram realizados diversos testes envolvendo diferentes métodos de seleção de atributos, transformações de bases e algoritmos de AM.

Resultados e Discussão

Para os testes sem transformação de bases, os melhores resultados encontrados estão dispostos na Tabela 1. Nesses casos, foi utilizado o método de seleção de atributos *CfsSubsetEval* combinado ao algoritmo *GreedyStepWise*. Os atributos (bandas espectrais) selecionados foram 1310 nm, 1520 nm e 1646 nm.

Tabela 1. Algoritmos com os melhores resultados em relação ao coeficiente de correlação (r), que é uma medida adimensional e Erro Absoluto Médio (MEA), expresso em gkg^{-1}.

Dados	Algoritmo	r	MEA (gkg^{-1})
Absorbância	SMOreg	0,93	0,0064
Refletância	IBk	0,98	0,0005
Transmitância	Random Forest	0,99	0,0014

Aplicando a transformação das bases multiplicando os valores por 100 e utilizando o método *Lasso Regression*, foram selecionados os atributos 1100 nm, 1522 nm, 1646 nm e 1648 nm. O melhor algoritmo encontrado foi o *Random Forest*, com profundidade máxima de 13 e máximo de iterações igual a 110. Nessas condições, o coeficiente de correlação foi de 0,94 e o erro absoluto médio foi de 2,8506 gkg^{-1}.

Conclusão

Os resultados apontam que os modelos são eficientes para a determinação do TOC, com alta correlação e baixo erro, demonstrando ser um método que pode ser usado como alternativa aos métodos tradicionais, facilitando as análises e dispensando a utilização de reagentes químicos poluentes.

Agradecimentos

À Fundação Araucária e à CAPES pelo apoio financeiro e ao C-LABMU/UEPG que disponibilizou a estrutura para a realização das análises experimentais.

ANGELOPOULOU, T. *et al*. From Laboratory to Proximal Sensing Spectroscopy for Soil Organic Carbon Estimation—A Review. **Sustainability**, v. 12, n. 2, p. 443–443, 7 jan. 2020.

BARRA, I. *et al*. Soil spectroscopy with the use of chemometrics, machine learning and pre-processing techniques in soil diagnosis: Recent advances–A review. **TrAC Trends in Analytical Chemistry**, v. 135, p. 116166, fev. 2021.

PAVAN, M.A. *et al*. **Manual de análise química do solo e controle de qualidade**. Londrina, Instituto Agronômico do Paraná, 1992. 38p.

ISSN 1676-0093