Mineração de dados espectrais do infravermelho próximo para estimativa do teor de nutrientes em folhas de trigo

Igor Guilherme Fidler (PIBIC/UEPG), Alessandro Stachak, Alaine Margarete Guimarães(Orientadora), e-mail: alainemg@uepg.br

Universidade Estadual de Ponta Grossa/Departamento de Informática

Área do conhecimento: 1.03.00.00-7 Ciência da Computação.
Subárea: 1.03.00-6 Metodologia e Técnicas da Computação.

Palavras-chave: Mineração de dados, Espectroscopia NIR, Trigo.

RESUMO

A análise de nutrientes em plantas é uma das tarefas fundamentais para a aplicação correta de insumos agrícolas. Os métodos tradicionais de análise de nutrientes em plantas geram custos ambientais e demandam de tempo para serem realizados. Como alternativa, tem-se a técnica de espectroscopia do infravermelho próximo - NIR. Neste trabalho foram analisadas 84 amostras de folha de trigo visando estimar os teores de Fósforo (P) e Enxofre (S). As amostras foram analisadas pelo espectrômetro NIR. Diferentes técnicas foram utilizadas para estimar o teor de nutrientes. Para a comparação entre as técnicas foram utilizados os coeficientes de correlação e de determinação, e a raiz quadrada do erro quadrático médio RMSE. O algoritmo com melhor desempenho, sem seleção prévia de atributos, foi a Máquina de Vetores de Suporte com R^2 de 0,87 e 0,59, para o fósforo e enxofre, respectivamente. Após a seleção de atributos, utilizando a técnica regressão "Ridge" obteve-se R^2 de 0,72 para o S, e com a regressão linear atingiu-se um R^2 de 0,88 para o P. Os resultados foram satisfatórios, quando comparados a literatura, demonstrando que a metodologia empregada foi eficiente.

INTRODUÇÃO

A estimativa de nutrientes de culturas na agricultura é importante para a aplicação correta de insumos melhorando a sua eficiência, contribuindo para a redução do impacto ambiental e aumento da produtividade no campo (TEIXEIRA, 2014; BOREM, 2015).

A espectroscopia do infravermelho próximo NIRS (do inglês *Near Infrared Spectroscopy*) é um método de análise rápido e não-destrutivo, que exige pouca preparação da amostra para a estimativa de materiais biológicos (PASQUINI, 2003). Essa técnica pode ser utilizada onde é necessário a análise de um grande número de amostras, e vem sendo empregada na agricultura e indústria alimentícia (CHODRAK, 2008). Porém, o uso do NIRS na estimativa de teor foliar de nutrientes ainda demanda de estudos para a sua efetiva adoção.

O NIRS por analisar uma grande faixa de espectro acaba por gerar dados de alta dimensionalidade. Assim, técnicas de redução de dimensionalidade são necessárias para que diferentes algoritmos possam ser utilizados, e os modelos gerados possam ser de mais fácil compreensão (Tan et al. 2006).

A Mineração de Dados (MD) é o processo da descoberta de informação útil em grandes repositórios de dados, para encontrar padrões e comportamentos que, de outra maneira, permaneceriam desconhecidos (Tan et al. 2006).

O objetivo deste trabalho foi verificar a eficiência do NIRS, aliado a algoritmos de mineração de dados, como alternativa para a estimativa do teor de Enxofre (S) e Fósforo (P) na folha de trigo.

MATERIAL E MÉTODOS

Foram utilizadas 84 amostras de folhas de trigo, as quais foram analisadas tanto no NIRS para obter a resposta espectral das mesmas, como também tais amostras foram previamente analisadas em laboratório para a obtenção do teor de S e P.

As 84 amostras foram lidas no equipamento NIRS, com três repetições, para identificar uma possível variação nos valores obtidos em cada análise. O espectrômetro utilizado foi o DA1650, com comprimento de onda variando de 1100 nm a 1648 nm. A leituras foram feitas com intervalos de 2 nm, resultando em 275 atributos, cada um correspondendo a um comprimento de onda, caracterizando uma base de dados de alta dimensionalidade.

A seleção de atributos, utilizando a técnica RFECV (Recursive Feature Elimination with Cross Validation), bem como a execução dos algoritmos de MD foram realizadas na linguagem Python 3.6 com o auxílio das bibliotecas SciKit-Learn 0.19.1 e Pandas 0.22.0.

Para a MD foram testados diferentes algoritmos de aprendizado de máquina, como a regressão linear com o método dos Mínimos Quadrados, a regressão linear "Ridge" e a Máquina de Vetores de Suporte (SVM) com os kernels linear, polinomial e RBF (Radial Basis Function). Para cada algoritmo, foi avaliada a eficiência do modelo gerado, antes e depois da técnica de redução de dimensionalidade ter sido aplicada.

A avaliação dos modelos de estimativa gerados foi feita pelos valores da: raiz quadrada do erro quadrático médio (RSME), coeficiente de correlação de Pearson (r) e o coeficiente de determinação (R^2).

RESULTADOS E DISCUSSÃO

A primeira parte do estudo consistiu na estimativa do teor de nutrientes utilizando todos os 275 atributos gerados pelo espectrômetro. A técnica que se destacou foi a SVM com os kernels RBF e polinomial, para a estimativa do P e S, respectivamente. Os melhores resultados podem ser observados na Tabela 1.

Tabela 1 – Melhores valores obtidos de r, R^2, RMSE obtidos sem a seleção de atributos, utilizando a técnica SVM.

Nutriente analisado	r	R^2	RMSE
P	0,94	0,87	0,32
S	0,86	0,59	0,54

Em seguida, por meio da seleção de atributos, foram selecionados 11 e 9 atributos para estimar o teor de S e P, respectivamente. Com esses novos subconjuntos de atributos, as técnicas que tiveram a melhor capacidade de estimar o teor de P e S foram a regressão linear com o método dos Mínimos Quadrados e a

regressão “Ridge”, respectivamente. Os melhores resultados estão apresentados na Tabela 2.

Tabela 2 – Melhores valores de r, R^2, RMSE obtidos após a seleção de atributos, utilizando a técnica de regressão linear.

Nutriente analisado	N° atributos	R	R^2	RMSE
P	9	0,95	0,88	0,31
S	11	0,90	0,72	0,47

Analisando os resultados das tabelas 1 e 2 foi possível verificar que a redução de atributos trouxe resultados semelhantes para a estimativa do teor de P. Para o S, a redução de atributos gerou uma melhora na capacidade de estimativa do nutriente.

O resultado obtido no presente trabalho utilizando nove atributos para a estimativa do teor de P foi superior ao de Xu et al (2017), que utilizaram sete atributos distintos e a técnica “PLSR” (Partial Least Square Regression). Para a estimativa do teor de S, os resultados obtidos neste trabalho foram semelhantes aos obtidos por Niu et al. (2014), que utilizaram a técnica “GA-PLSR” (Genetic Algorithm Partial Least Square Regression).

Conclusão

A espectroscopia NIR aliada a técnicas de mineração de dados apresentou boa capacidade de estimar o teor de S e P presentes nas folhas de trigo.

Para a base de dados completa, de alta dimensionalidade, a técnica de SVM foi a mais adequada, enquanto que após a seleção de atributos, algoritmos de regressão linear foram os que geraram os melhores resultados.

A redução de dimensionalidade possibilitou a melhora dos resultados para a estimativa de S, mas não afetou a estimativa de P.

Agradecimentos

Ao Laboratório de Fertilidade do Solo da UEPG por fornecer as amostras e resultados da análise laboratorial e ao C-LabMu da UEPG por disponibilizar o espectrômetro utilizado na pesquisa.

Referências

BOREM, P. L. S. A. Trigo: do plantio à colheita. Viçosa, Minas Gerais: Ed. UFV, 2015.

CHODAK, M. Application of near infrared spectroscopy for analysis of soils, litter and plant materials. Polish Journal Of Environmental Studies, 2008, vol. 17(5), p. 631-642, 2008.

NIU, W; et al. Chemical Composition and Calorific Value Prediction of Wheat Straw at Different Maturity Stages Using Near-Infrared Reflectance Spectroscopy. Energy Fuels, v. 28, p. 7474-7482, 2014.

PASQUINI, C. Near Infrared Spectroscopy: Fundamentals, Practical Aspects and Analytical Applications. Journal of the Brazilian Chemical society, São Paulo, v. 14, n, 2, p. 198-219, 2003.

TAN, P.; et al. Introduction to data mining. Pearson education, 2006.

TEIXEIRA, S. Determinação de modelo de estimativa de teores de carbono em solos utilizando máquina de vetores de suporte e refletância espectral. 2014. 64 f. Dissertação (Mestrado em Computação Aplicada) – Setor de Ciências Agrárias e de Tecnologia, Universidade Estadual de Ponta Grossa, Ponta Grossa, 2014.

XU, S; et al. Rapid determination of carbon, nitrogen, and phosphorus contents of field crops in China using visible and near-infrared reflectance spectroscopy. Crop Science, January-February, vol. 57(1), p. 475-489, 2017.