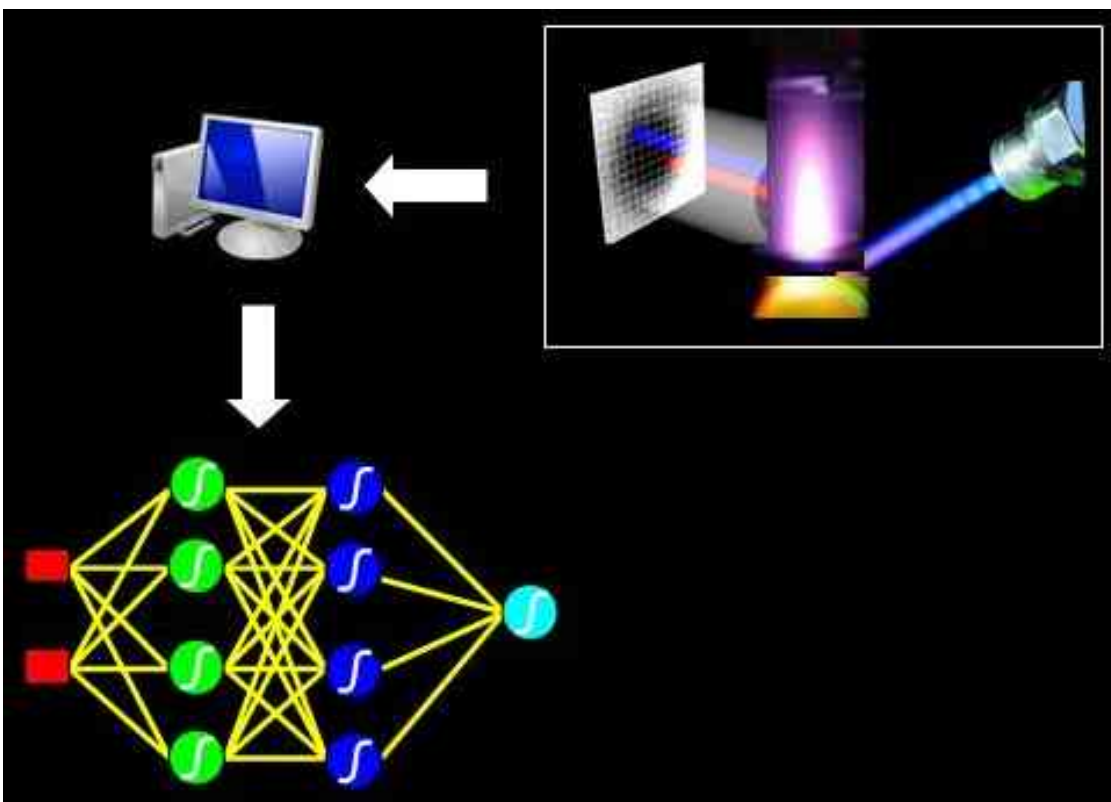


**Métodos de Seleção de
Comprimentos de Onda e Redes
Neurais Artificiais como Estratégias
de Calibração em Espectroscopia de
Emissão Ótica Induzida por Laser**



ISSN 1678-0434

Setembro, 2007

*Empresa Brasileira de Pesquisa Agropecuária
Embrapa Instrumentação Agropecuária
Ministério da Agricultura e do Abastecimento*

Boletim de Pesquisa e Desenvolvimento 18

Métodos de Seleção de Comprimentos de Onda e Redes Neurais Artificiais como Estratégias de Calibração em Espectroscopia de Emissão Ótica Induzida por Laser

Edilene Cristina Ferreira
Ednaldo José Ferreira
Débora Marcondes Bastos Pereira Milori
Robson Marcel da Silva
Ladislau Martin Neto

Embrapa Instrumentação Agropecuária
São Carlos, SP
2007

Exemplares desta publicação podem ser adquiridos na:

Embrapa Instrumentação Agropecuária

Rua XV de Novembro, 1452
Caixa Postal 741
CEP 13560-970 - São Carlos-SP
Fone: (16) 3374 2477
Fax: (16) 3372 5958
www.cnpdia.embrapa.br
E-mail: sac@cnpdia.embrapa.br

Comitê de Publicações da Unidade

Presidente: Dr. Carlos Manoel Pedro Vaz
Membros: Dra. Débora Marcondes Bastos Pereira Milori,
Dr. João de Mendonça Naime,
Dr. Washington Luiz de Barros Melo
Valéria de Fátima Cardoso
Membro Suplente: Dr. Paulo Sérgio de Paula Herrmann Junior

Supervisor editorial: Dr. Victor Bertucci Neto
Normalização bibliográfica: Valéria de Fátima Cardoso
Tratamento de ilustrações: Valentim Monzane
Capa foto montagem: Odílio B. Garrido Assis
Editoração eletrônica: Valentim Monzane

1ª edição

1ª impressão (2007): tiragem 300

Todos os direitos reservados.

**A reprodução não-autorizada desta publicação, no todo ou em parte,
constitui violação dos direitos autorais (Lei nº 9.610).**

**CIP-Brasil. Catalogação-na-publicação.
Embrapa Instrumentação Agropecuária**

F383m Ferreira, Edilene Cristina

Métodos de seleção de comprimentos de onda e redes neurais artificiais como estratégias de calibração em espectroscopia de emissão ótica induzida por laser. / Edilene Cristina Ferreira, Ednaldo José Ferreira, Débora Marcondes Bastos Pereira Milori, Robson Marcel da Silva, Ladislau Martin-Neto. - São Carlos: Embrapa Instrumentação Agropecuária, 2007.

11 p. - (Embrapa Instrumentação Agropecuária. Boletim de Pesquisa e Desenvolvimento, ISSN 1678-0434; 18)

1. LIBS. 2. Análise de solos. 3. Estratégia para calibração. 4. RNA.I. Ferreira, Ednaldo José. II. Milori, Débora Marcondes Bastos Pereira, III. Silva, Robson Marcel da. IV. Martin-Neto, Ladislau. V. Título. VI. Série.

Sumário

Resumo	4
Abstract	5
Introdução	6
Materiais e Métodos	7
Resultados e Discussão	8
Conclusões	10
Referências	11

Métodos de Seleção de Comprimentos de Onda e Redes Neurais Artificiais como Estratégias de Calibração em Espectroscopia de Emissão Ótica Induzida por Laser

Edilene Cristina Ferreira¹

Ednaldo José Ferreira²

Débora Marcondes Bastos Pereira Milori³

Robson Marcel da Silva⁴

Ladislau Martin Neto⁵

Resumo

Espectroscopia de emissão ótica induzida por laser (LIBS, do acrônimo Laser Induced Breakdown Spectroscopy) é uma técnica analítica avançada utilizada para análise elementar, cuja medida baseia-se em determinações diretas do sinal de emissão óptica de espécies excitadas em um plasma indutivamente gerado por pulso de laser. LIBS apresenta um grande potencial para efetuar análise direta, independentemente do estado físico da amostra (sólido, líquido ou gás). Atualmente LIBS tem sido empregado facilmente em análises qualitativas, contudo análises quantitativas ainda têm sido uma dificuldade, pois a calibração analítica é ainda um ponto bastante questionável. Rede neural artificial (RNA) é um paradigma de aprendizado de máquina inspirado em sistemas nervosos biológicos. Recentemente RNAs têm sido utilizadas em muitas aplicações e sua capacidade de classificação e predição são especialmente importantes para análise espectral. Neste trabalho duas estratégias foram utilizadas para selecionar comprimentos de onda importantes para determinação de Cu em amostras de solo. Regressão linear simples e abordagem Wrapper foram as estratégias empregadas para selecionar um conjunto de comprimentos de onda. Os comprimentos de onda selecionados foram então utilizados para o treinamento de uma RNA, com o objetivo de utilizá-la como estratégia de calibração de LIBS. Um conjunto heterogêneo com 59 amostras certificadas de diferentes tipos de solos foram submetidos a análise por LIBS e seus espectros foram utilizados para calibração.. A estratégia de validação cruzada foi adotada para determinação da exatidão do método. A RNA gerada apresentou um bom coeficiente de correlação ($r = 0,96$) para a predição de Cu nas amostras, apesar da baixa resolução espectral do equipamento utilizado. O método proposto apresentou limite de detecção (LD) igual a 2.26 mg dm^{-3} .

Termos para indexação: LIBS, análise de solos, estratégia para calibração, RNA

¹Química, Dra., Embrapa Instrumentação Agropecuária, C.P. 741, CEP 13560-970, São Carlos, SP, edilene@cnpdia.embrapa.br

²Ciência da Computação, MSc., Embrapa Instrumentação Agropecuária, ednaldo@cnpdia.embrapa.br

³Física, Dra., Embrapa Instrumentação Agropecuária, debora@cnpdia.embrapa.br

⁴Químico, Graduado, Instituto de Química de São Carlos, IQSC - USP, 400, C.P. 780, São Carlos, SP, CEP 13560-970

⁵Físico, Dr., Embrapa Instrumentação Agropecuária, martin@cnpdia.embrapa.br

Wavelength Selection Methods and Artificial Neural Networks as Calibration Strategy in Laser Induced Breakdown Spectroscopy (LIBS)

Edilene Cristina Ferreira¹

Ednaldo José Ferreira²

Débora Marcondes Bastos Pereira Milori³

Robson Marcel da Silva⁴

Ladislau Martin Neto⁵

Abstract

Laser Induced Breakdown Spectroscopy (LIBS) is an advanced analytical technique for elemental determination based on the direct measurement of optical emission of excited species on a laser induced plasma. LIBS technique has great potential to accomplish direct analysis independently of physical sample state (solid, liquid or gas). Presently, LIBS has been easily employed for qualitative analysis, nevertheless, in order to perform quantitative analysis, some effort is still required since calibration represents a difficult issue in the development of LIBS methods. Artificial Neural Network (ANN) is a machine learning paradigm inspired on biological nervous systems. Recently, ANNs have been used in many applications and its classification and prediction capabilities are especially useful for spectral analysis. In the present work, two strategies: simple linear regression (SLR) and wrapper approach were employed to select a set of important wavelengths for Cu prediction in soil samples. The wavelengths selected were used to training an ANN, which was used as LIBS calibration strategy. Spectra of 59 samples from a heterogenic set of certified soil samples and their certified Cu concentration were used for calibration. Cross validation was applied, following ANN training, for verification of prediction accuracy. The ANN showed good efficiency ($r=0.96$) for prediction of Cu amounts, although low spectral resolution of the employed detector. The proposed method presented 2.26 mg dm^{-3} of limit of detection (LOD).

Keywords: LIBS, soil analysis, strategy calibration, ANN

Introdução

Espectrometria de emissão óptica com plasma induzido por laser é uma técnica espectroanalítica que utiliza um plasma gerado por pulsos de laser de alta energia para preparar a amostra e excitar os analitos em um único passo. Essa técnica tem possibilitado medidas elementares de forma não intrusiva em qualquer tipo de material, independente de seu estado físico, sólido, líquido ou gasoso (SANTOS JUNIOR et al., 2006).

A focalização do laser em uma pequena área da superfície da amostra provoca a ablação de pequena quantidade do material, da ordem de nanogramas a picogramas. O material ablado sofre um superaquecimento instantâneo e as colisões entre elétrons livres e elétrons ligados aos átomos resultam em uma avalanche eletrônica provocando a geração de um plasma com temperaturas na faixa de 10.000 a 20.000 K (MARKOLF, 1999). A temperatura do plasma, que permanece aproximadamente constante por um curto espaço de tempo (< 10 s), é suficientemente energética para excitar os elétrons das espécies atômicas e iônicas presentes no plasma. Com o resfriamento do plasma, os elétrons excitados retornam aos seus estados fundamentais liberando a energia absorvida na forma de radiação eletromagnética, cujas frequências são características de cada espécie excitada. A identificação dos elementos presentes na amostra se dá pela identificação das linhas de emissão, que já são bem conhecidas, e as intensidades relativas das linhas de emissão podem ser utilizadas para quantificação da concentração das espécies presentes na amostra (YUEH et al., 2000; FERIOLI e BUCKLEY, 2006). O uso de curvas de calibração baseadas na intensidade absoluta é um método estatístico bastante utilizado em química analítica para determinação de diferentes analitos (SKOOG et al., 1996). Contudo, a utilização desse tipo de calibração é totalmente dependente do conhecimento da composição exata da amostra e no caso de LIBS, onde o grande ganho analítico é a possibilidade de determinação direta de uma amostra, a reprodução de uma curva de calibração com características semelhantes às das amostras torna-se quase impossível. Além disso, o efeito de matriz é mais acentuado no plasma de LIBS que em outros plasmas analíticos, devido a alta densidade do material da amostra no plasma (CREMERS e RADZIEMSKI, 2006). Por isso, a sensibilidade aos efeitos de matriz e sua influência sobre os resultados é ainda um problema não resolvido para a técnica LIBS (ANZANO et al., 2006).

Redes Neurais Artificiais são modelos capazes de solucionar tarefas de classificação e predição de padrões. As RNAs resolvem problemas complexos por meio da imitação da estrutura e do funcionamento do cérebro humano (BRAGA et al., 2000). A utilização de uma RNA na solução de uma tarefa inicia-se por uma fase de aprendizagem, que é quando a rede extrai informações relevantes de padrões de informação apresentados a ela e cria uma representação matemática própria para o problema (BRAGA et al., 2000).

Na fase de treinamento o conhecimento é induzido nos pesos das conexões por meio de um algoritmo de aprendizado que consiste em um conjunto de procedimentos bem definidos para adaptar os parâmetros de uma RNA com o objetivo de aprender alguma função específica. Os dados de treinamento fornecem os exemplos de referência e o algoritmo de aprendizado incorpora o conhecimento na RNA (FERREIRA, 2005).

A fase de validação, subsequente à fase de treinamento, tem o objetivo de verificar a capacidade de generalização da rede. Espera-se uma rede com capacidade de generalização adequada que seja capaz de categorizar os dados de teste corretamente (FERREIRA, 2005).

No presente trabalho, uma RNA do tipo MultiLayer Perceptron (MLP) foi utilizada como método de calibração para regressão e estimativa do teor de cobre em solos a partir dos espectros obtidos por um equipamento LIBS portátil.

Materiais e Métodos

Um conjunto de 59 amostras de diferentes solos brasileiros com teores certificados pelo Instituto Agrônomo de Campinas (IAC) para alguns elementos foi utilizado para construção do conjunto de treinamento/calibração. Somente o teor do elemento Cu foi considerado, sendo que as amostras selecionadas apresentavam teores variados desse elemento desde concentrações de 0,2 mg dm⁻³ à 10,4 mg dm⁻³.

Antes da análise, as amostras foram submetidas à moagem criogênica para redução da heterogeneidade. Após a moagem, as amostras foram submetidas a uma pressão de 15 ton em recipiente de aço, para obtenção de pastilhas apropriadas ao porta amostras do sistema LIBS.

Os espectros das amostras foram obtidos utilizando-se um equipamento LIBS portátil adquirido da Stellar Net. O sistema é composto por um laser NdYAG (1064 nm), que opera com pulsos simples a potência de 25 mJ, uma câmara de amostragem e quatro espectrômetros CCD (Charge Coupled Device) modelo Epp 2000 com resolução espectral de 0,2 nm, que juntos fazem aquisição de dados na faixa espectral de 190 nm a 1000 nm. O tempo entre o pulso do laser e a aquisição dos espectros foi de 2 s. A aquisição dos dados foi efetuada por um programa computacional desenvolvido pela Stellar Net.

Para melhorar a exatidão dos resultados foram capturados 60 espectros de cada amostra sendo cada espectro correspondente a diferentes regiões da pastilha e um disparo foi utilizado para limpeza da amostra antes da captura dos espectros. Os dados espectrais correspondentes a faixa de 190 nm a 400 nm foram submetidos a correção da linha base e a cálculos de média.

A RNA adotada foi a MLP, desenvolvida em JAVA no sistema Weka que disponibiliza funções para treinamento de RNAs (WITTEN e FRANK, 2005). Técnicas de descoberta de conhecimento KDD (do inglês *Knowledge Discovery in Databases*) foram utilizadas especialmente na etapa de pré-processamento dos dados espectrais. Duas estratégias para Kohavi e John (1997). Após a seleção dos comprimentos de ondas relevantes para a tarefa de aprendizado foram adotadas: (1) ajuste de um modelo por regressão linear simples e (2) aplicação da abordagem *Wrapper* (tarefa de seleção dos comprimentos de ondas, a MLP foi treinada com o algoritmo *backpropagation*. Os parâmetros do *backpropagation*, assim como da topologia da MLP, foram os seguintes: Taxa de aprendizado = 0,1; Momento = 0,2; Número de épocas = 3300 e duas camadas intermediárias com $t + 1$ elementos (neurônios), onde t foi o número de comprimentos de ondas selecionados pela estratégia referida.

O desempenho da MLP foi avaliado em função do maior coeficiente de correlação (Pearson) e do menor erro médio absoluto (EMA) calculado em função do conjunto de dados. A validação cruzada em 10 dobras foi o método adotado para validar a rede. Com a utilização dessa técnica foi possível estimar a precisão do método.

Resultados e Discussão

A calibração analítica é um ponto essencial para a qualidade da análise por isso, o primeiro passo a ser definido na proposição de um novo método analítico é o material e o método mais adequado para a calibração (KURFÜRST, 1998). No caso da análise direta de solos por LIBS, as linhas de emissão dos elementos são fortemente influenciadas pelo efeito da matriz, uma vez que a matriz do solo é bastante complexa. Devido à dificuldade em se reproduzir a matriz do solo, a construção de uma curva de calibração convencional é uma tarefa praticamente impossível. Além das dificuldades encontradas para calibração, a utilização de um equipamento de baixa resolução espectral (0,2 nm) também dificultou o isolamento da linha de interesse e conseqüentemente a identificação de sua intensidade.

Os métodos de calibração multivariados apresentam características bastante interessantes que podem ajudar a minimizar tanto os problemas relacionados aos efeitos de matriz como os problemas relativos à resolução espectral do equipamento. As RNAs são métodos multivariados, cujas características permitem que elas sejam aplicadas para funcionar como calibração de um método analítico. A diferença básica entre os métodos de calibração convencionais e as RNAs reside no fato de que as RNAs geram determinadas funções matemáticas que normalmente não são lineares. Além disso, uma RNA pode extrair informações diferentes da intensidade de pico se essas informações forem “julgadas” relevantes para estimar a variável de interesse. Por esse motivo o efeito de matriz e a resolução espectral podem ser minimizados por meio de uma método de calibração baseado em um modelo de RNA.

Antes do processo de treinamento e aprendizagem da RNA foram utilizadas duas estratégias para seleção de comprimentos de onda relevantes para a predição da concentração de cobre nas amostras. A primeira estratégia, empiricamente embasada, consistiu em ajustar um modelo de regressão linear simples para cada comprimento de onda. O objetivo dessa estratégia foi identificar o comprimento de onda de maior correlação linear com as concentrações de cobre presentes nas amostras. O melhor modelo de regressão linear ajustado individualmente a cada comprimento de onda, usando como referência todos os dados para calibração, resultou no comprimento 224,47 nm, cujo erro médio absoluto e o coeficiente de correlação de Pearson calculados entre as concentrações e a predição do modelo foram 3,11 e 0,73, respectivamente. É interessante observar que esse comprimento de onda (considerando-se os deslocamentos experimentais) corresponde ao comprimento de onda da linha de emissão atômica do cobre descrita na literatura (Cu I, 224,426 nm).

A segunda estratégia para seleção consistiu da aplicação da abordagem *Wrapper* (KOHAVI e JOHN, 1997) para seleção de um subconjunto de maior relevância para a tarefa de aprendizado da RNA. O algoritmo de aprendizado baseado em instâncias - *Instance-based learning* (AHA et al., 1991) foi agregado à abordagem *Wrapper*. O IBk (AHA et al., 1991), vertente do *Instance-based* que estima a saída, baseado na média dos k-vizinhos mais próximos foi utilizado por apresentar um desempenho computacional significativamente superior ao algoritmo *backpropagation* (RUMMELHART et al., 1986) viabilizando, assim, a tarefa de seleção dos comprimentos relevantes. A média entre os 5 vizinhos foi previamente definida ($k = 5$) e dessa forma, foi obtido o subconjunto {207,14 nm, 221,17 nm, 247,28 nm, 270,09 nm, 398,97 nm}. Desse subconjunto somente 270,09 nm coincide com uma linha iônica de cobre, os demais provavelmente correspondem a outros elementos que contribuem, de alguma forma, para estimar a concentração de cobre. Essa propriedade identificada durante o processo seleção poderá ser utilizada para avaliar as relações interelementares que ocorrem no solo.

Após o treinamento, a RNA foi validada com 10 dobras mutuamente exclusiva para certificar o desempenho em dados distintos dos dados de calibração. Para cada dobra, 53 amostras foram utilizadas no treinamento da RNA e as 6 restantes na validação. Os resultados obtidos com os dados de validação para as 10 dobras são sumariados na Fig. 1. Conforme podem ser observado, os valores preditos pela RNA apresentaram boa correlação com os valores certificados, sendo o coeficiente de correlação de Pearson calculado para todas as amostras nas 10 dobras de 0,96.

O limite de detecção calculado de acordo com Sirven e colaboradores (SIRVEN et al., 2006) foi de $2,26 \text{ mg dm}^{-3}$, o qual pode ser considerado bom devido a baixa resolução espectral do equipamento utilizado e o grande efeito de matriz observado nas amostras de solos.

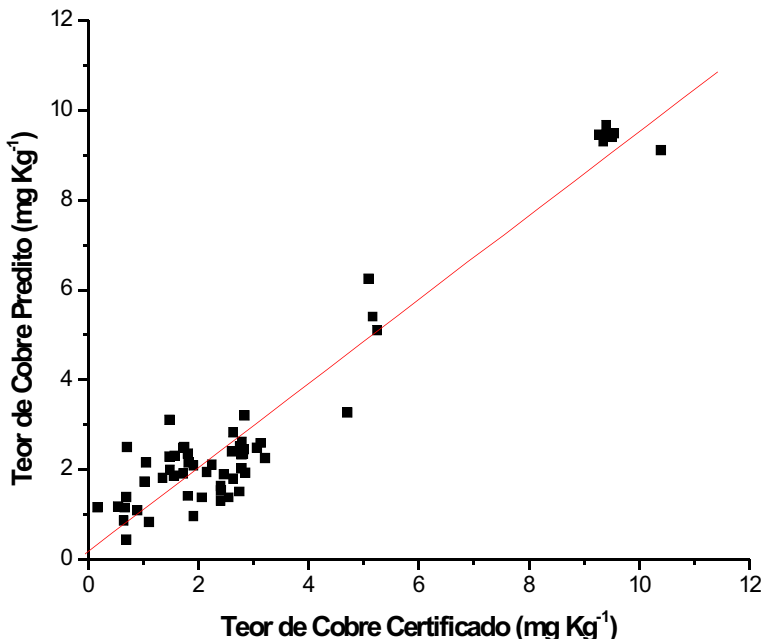


Fig. 1 - Correlação entre os teores de cobre preditos pela RNA e os teores certificados nas amostras de solos utilizadas

Conclusões

A dificuldade em se estabelecer uma eficiente calibração para métodos analíticos é significativamente aumentada quando se trabalha com análise direta de sólidos devido à alta complexidade das matrizes. A complexidade dos espectros de linhas de emissão obtidas por sistemas LIBS de baixa sensibilidade e baixa resolução espectral dificulta o isolamento visual das linhas de interesse. A utilização de RNA para a função de calibração associada a estratégias para seleção de comprimentos de onda mostrou-se bastante eficiente.

O comprimento de onda selecionado como o mais relevante pela estratégia de seleção baseada em regressão linear simples foi uma linha de emissão atômica de cobre, selecionada sem prévia informação e de difícil visualização pela simples observação dos espectros. As linhas de emissão selecionada por meio da estratégia *Wrapper* contribuem para o aumento do desempenho da RNA e ainda podem apresentar importantes informações sobre as relações interelementares no solo. A RNA desenvolvida para essa estratégia mostrou um excelente desempenho, com um bom coeficiente de correlação de 0,96. O limite de detecção do método de 2.26 mg dm^{-3} foi considerado bom baseado em comparações aos apresentados na literatura.

Referências

- AHA, D.; KIBLER, D.; ALBERT, M. K. Instance-based learning algorithms. Machine Learning, Boston, v. 6, p. 37-66, 1991.
- ANZANO, J. M.; VILLORIA, M. A.; RUÍZ-MEDINA, A.; LASHERAS, R. J. Laser-Induced Breakdown Spectroscopy for Quantitative Spectrochemical Analysis of Geological Materials: Effects of the Matriz and Simultaneous Determination. Analytica Chimica Acta, Amsterdam, v. 575, p. 230-235, 2006.
- BRAGA, A. P.; LUDERMIR, T. B.; CARVALHO, A. C. P. L. F. Redes Neurais Artificiais: Teoria e Aplicações. Rio de Janeiro: LTC, 2000.
- CREMERS, D. A.; RADZIEMSKI, L. J. Handbook of Laser-Induced Breakdown Spectroscopy, England: Wiley, 2006.
- FERIOLI, F.; BUCKLEY, S. G. Measurements of hydrocarbons using laser-induced breakdown spectroscopy. Combustion and Flame, New York, v. 144, n. 3p. 435-447, 2006.
- FERREIRA E. J. 2005. Abordagem Genética para seleção de um conjunto reduzido de características para construção de ensembles de redes neurais: Aplicação a língua eletrônica. Dissertação (Mestrado) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos.
- KOHAVI, R.; JOHN, G. H. Wrappers for feature subset selection. Artificial Intelligence, Amsterdam, v. 97, p. 273-324, 1997.
- KURFÜRST, U. Solid Sample Analysis. [S. l.]: Springer, 1998.
- MARKOLF, H. N. Laser-Tissue Interactions. 1st ed. Berlin: Springer, 1999.
- RUMMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. Nature, London, v. 323, p. 533-536, 1986.
- SANTOS-JÚNIOR, D.; TARELHO, L. V. G.; KRUG, F. J.; MILORI, D. M. B. P.; MARTIN-NETO, L. ; VIEIRA-JÚNIOR, N. D. Espectrometria de emissão óptica com plasma induzido por laser (LIBS): fundamentos, aplicações e perspectivas. Analytica, São Paulo, v. 5, n. 24, p. 72-81, 2006.
- SIRVEN, J. -B.; BOUSQUET, B.; CANIONI, L.; SARGER, L. Laser-Induced Breakdown Spectroscopy of composite samples: comparison of advanced chemometrics methods. Analytical Chemistry, Washington, v. 78, p. 1462-1469, 2006.

SKOOG, D. A.; WEST, D. M.; HOLLER, F. J. Fundamentals of Analytical Chemistry. 7th ed. Orlando: Saunders College Publishing, 1996.

WITTEN, I. H.; FRANK, E. Data Mining: Practical machine learning tools and techniques. 2nd ed. San Francisco: Elsevier, 2005.

YUEH, F. Y.; SINGH, J. P.; ZHANG, H. Laser-induced Breakdown Spectroscopy, Elemental Analysis. In: MEYERS, R. A. (Ed.). Encyclopedia of Analytical Chemistry. Chichester, UK: John Wiley and Sons, 2000. p. 2066-2087.



Empresa Brasileira de Pesquisa Agropecuária

Embrapa Instrumentação Agropecuária

Ministério da Agricultura, Pecuária e Abastecimento

Rua XV de Novembro, 1452 - Caixa Postal 741 - CEP 13560-970 - São Carlos - SP

Telefone: (16) 3374 2477 - Fax: (16) 3372 5958

www.cnpdia.embrapa.br - sac@cnpdia.embrapa.br

**Ministério da
Agricultura, Pecuária
e Abastecimento**

