



MELHORIAS NA ANÁLISE DISCRIMINATÓRIA DE CAFÉS COMERCIAIS BRASILEIROS APLICANDO O ALGORITMO DE SELEÇÃO DE VARIÁVEIS OPS[®] SOBRE DADOS CROMATOGRÁFICOS

Juliano S. Ribeiro (PG)^{1,2*}, Reinaldo F. Teófilo (PG)¹, João P. A. Martins (PG)¹ e Márcia M. C. Ferreira (PQ)¹

jribeiro@iqm.unicamp.br

1. Universidade Estadual de Campinas (UNICAMP), Instituto de Química: Laboratório de Quimiometria Teórica e Aplicada (LQTA);

2. Instituto Agrônomo de Campinas (IAC): Centro de Análise e Pesquisa Tecnológica do Agronegócio do Café "Alcides Carvalho" – Campinas – SP;

Palavras Chave: Seleção de variáveis, PLSDA, café torrado, OPS

Introdução

Atualmente, devido aos recentes avanços computacionais, uma grande quantidade de dados gerada por diferentes instrumentos analíticos vem sendo tratada com sucesso por diversas ferramentas quimiométricas. Dentre essas ferramentas, PLSDA (Partial Least Square – Discriminant Analysis) vem sendo utilizada com o intuito de encontrar variáveis que possam discriminar diferentes classes [1]. Em muitos casos, o número de variáveis geradas pelos instrumentos é relativamente grande e apenas algumas são realmente importantes para expressar certas características ou discriminar classes. Deste modo, é crescente o interesse na seleção destas variáveis que apresentam um alto grau de importância na construção de modelos. Seguindo nesta linha, o algoritmo de seleção de variáveis OPS [2] (Ordered Predictor Selection) aparece como uma alternativa nova e de grande importância neste processo. Este algoritmo consiste em três etapas: (1) obtenção de um vetor informativo; (2) ordenamento decrescente das variáveis por este vetor; (3) investigação das variáveis ordenadas. O objetivo deste trabalho é demonstrar o potencial do algoritmo OPS para selecionar variáveis em análises discriminantes.

Experimental

Para este trabalho, 14 amostras de cafés comerciais brasileiros de uma mesma marca foram utilizadas, sendo sete delas extra-fortes e as outras sete tradicionais. A técnica analítica de extração utilizada foi a microextração em fase sólida (SPME) [3]. Cada amostra foi analisada em triplicata. Os dados cromatográficos foram organizados em uma matriz \mathbf{X} (42x7500). Para o

alinhamento dos cromatogramas foi utilizado o algoritmo COW [4] e para seleção de 10 amostras para previsão foi utilizado o algoritmo de Kennard e Stone [5].

Resultados e Discussão

O modelo de calibração realizado com os dados sem seleção de variáveis utilizando 3 variáveis latentes (VL) apresentou erro padrão de validação cruzada (RMSECV) de 0,2098 e coeficiente de correlação de validação cruzada (r_{cv}) de 0,932. Na validação cruzada 2 amostras foram classificadas incorretamente. Na previsão mais duas amostras foram classificadas em classes erradas. Utilizando o algoritmo OPS foram selecionadas 800 variáveis das 7500. O modelo de calibração com 3 VL apresentou RMSECV = 0,1288 e r_{vc} = 0,986. Neste modelo, tanto na calibração quanto na previsão todas as amostras foram classificadas corretamente.

Conclusões

A partir dos resultados obtidos em ambos os modelos, fica claro que os dados tratados com o algoritmo de seleção de variáveis OPS apresentam melhores resultados. Outros dados também vêm sendo tratados com a mesma metodologia e apresentam resultados excelentes.

Agradecimentos

A Capes e ao CNPq pelo auxílio financeiro.

- [1] Barker, M., Rayens, W., J. Chem., 17 (2003) 166.
- [2] Teófilo, R., et al., 10th Inter. Conf. Chem. Anal. Chem. (2006).
- [3] Bicchi, C., et al., J. Agric. Food Chem., 50 (2002) 449.
- [4] Nielsen, N. P. V., et al., J. Chrom. A, 805 (1998) 17.
- [5] Kennard, R. W., and Stone, L. A., Technometrics, 11 (1969) 137.