DESENVOLVIMENTO DE UM PROGRAMA COMPUTACIONAL PARA CALIBRAÇÃO MULTIVARIADA BASEADO NO MÉTODO DOS MÍNIMOS QUADRADOS PARCIAIS

Ronei J. Poppi e Célio Pasquini - Instituto de Química-UNICAMP C.P. 6154 - CEP 13081 - Campinas (SP).

ABSTRACT :

DEVELOPMENT OF A MULTIVARIATE CALIBRATION COMPUTER PROGRAM BASEAD ON PARTIAL LEAST SQUARES METHOD (PLS).

A computer program to perform multivariate calibration based on a partial least squares method (PLS) has been developed. This program was written in Quick Basic 4.5 for IBM PCs and had been successfully tested in cromatographic dates with the same performance of comercial packages.

INTRODUÇÃO :

Nos últimos anos, métodos de calibração multivariada tem sido cada vez mais utilizados em química analítica, e dentre estes, há um enorme destaque ao método dos mínimos quadrados parciais (PLS - partial least squares) devido ao seu excelente desempenho e facilidade de programação. Esse método de calibração multivariada encontra-se disponível apenas em pacotes computacionais comerciais de quimiometria, limitando, assim, o acesso daqueles que desejam utilizá-lo. Para tornar o método mais conhecido e estender sua utilização a um número maior de pessoas, decidiu-se desenvolver um programa computacional em linguagem Quick Basic 4.5 para micro-computadores tipo IBM - PC, tendo como base o trabalho de Geladi e Kowalski[1] que descreve todos os aspectos mais importantes da metodologia.

O programa foi testado em um conjunto de dados cromatográficos[2] onde o método PLS contido no pacote computacional SIMCA 3B[3] já havia sido utilizado com sucesso.

O MÉTODO DOS MÍNIMOS QUADRADOS PARCIAIS (PLS) :

É possível estabelecer uma relação entre duas matrizes de dados X e Y, quando houver uma dependência entre as propriedades que descrevem cada uma delas. A forma de estabelecer esta relação é a base da calibração multivariada.

O cálculo dos componentes principais de uma matriz de dados utilizando-se o algoritmo NIPALS[4] (Nonlinear Iterative Partial Least Squares) é a base do método dos mínimos quadrados parciais (PLS). Pela análise de componentes principais, uma matriz X é decomposta em :

$$X=TB$$

onde B é a matriz de autovetores de X'X e T é a correspondente matriz de *scores*.

No PLS tanto a matriz das variáveis independentes (X) como a das variáveis dependentes (Y) são representadas por seus *scores*, pela modelagem de componentes principais :

$$X = TP \text{ e } Y = UQ.$$

Uma relação entre os dois blocos pode ser feita correlacionando-se os *scores* do bloco Y (u), com os *scores* do bloco X (t), para cada componente principal de cada vez. Um modelo linear é utilizado para esta relação :

$$u=bt.$$

Esse modelo no PLS é ainda otimizado, com as informações dos dois blocos sendo manipuladas simultaneamente para que se obtenha a melhor correlação possível.

O algoritmo utilizado para o desenvolvimento do programa foi o seguinte :

- FASE DE CALIBRAÇÃO:

- X E Y são centralizados e escalonados para variância unitária

- índice de h = 1.

(1) u_{inic} = algum y_i e t_{inic} algum x_i

- para o bloco X : (2) w' = u'X/u'u (pesos de X).
(3) Normalize w para || w || = 1.
(4) t = X w/w'w (scores de X).

- para o bloco Y : (5) q' = t'Y/t't (loadings de Y).
(6) Normalize q para || q || = 1.
(7) u = Y q/q'q (scores de Y).

- cheque convergência : compare t do passo (4) com o da interação anterior. Se $|| t - t_{ant.} || > 10^{-6} || t ||$ volte ao passo (2), então siga no passo (9).

(9) p' = t'X/t't (loadings de X)
(10) Normalize p para || p || = 1.

- encontre o coeficiente b (relação entre t e u) :

(11) b = u't/t't

- cálculo dos resíduos :

Eh = Eh-1 - thph' ; X = Eo
Fh = Fh-1 - bhthqh' ; Y = Eo

- incremente h, substitua Eh por X e Fh por Y, volte ao passo (1) para implementação do próximo componente principal.

- FASE DE PREVISÃO :

- centralize e escalone Xp (matriz de previsão) com os dados da calibração.

(1) índice de h = 1 (Xp = Eo).
(2) t = Eh-1 wh (scores de Xp)
(3) Eh = Eh-1 - thph (resíduos de Xp)

- para o bloco Y : (4) Fh = $\sum$ bhthqh' (valores previstos)

- incremente h , volte ao passo (2) e repita o procedimento até o número desejado de componentes principais. (note que p, q, w, b, são da fase de calibração).

- VALORES PREVISTOS : $Yp = \sum_{h=1}^{a} Fh$ (a = num. de componentes)

TESTE DO PROGRAMA :

O programa foi testado em determinações simultâneas em cromatografia gasosa onde havia superposição dos picos. O conjunto de dados utilizado constituiu-se de 41 variáveis independentes (os cromatogramas digitalizados) e 3 variáveis independentes (as concentrações de misturas de Etanol, Isoctano e Tolueno). Para maiores informações ver ref. 2.

Os resultados obtidos na análise dos dados cromatográficos mostraram que o programa foi desenvolvido de maneira satisfatória, apresentando resultados idênticos quando comparados aos obtidos com o pacote computacional SIMCA 3B.

AGADECIMENTOS :

Os autores agradecem à FAPESP (Proc. 90/4168-0) pelo apoio financeiro.

BIBLIOGRAFIA :

1. Geladi,P., Kowalski,B.R., Anal. Chim. Acta (1986), 185, 1.
2. Poppi,R.J., Faigle,J.F.G., Scarminio,I.S., Bruns,R. E, J. of Chromatogr. (1991), 539, 123.
3. Principal Data Components, 2505 Shepard Blvd., Columbia, MO 65201, USA.
4. Vandeginste, B. G. M., Sielhorst, C., Gerritsen, M., Trends Anal. Chem. (1988), 7(8), 286.