

Desenvolvimento de um algoritmo para correção de dados instrumentais: aplicação em alinhamento de cromatogramas

Guilherme P. Sabin (PG)*, Paulo H. Março (PQ), Fabio Augusto (PQ), Ronei J. Poppi (PQ).
gpsabin@gmail.com

UNICAMP - Instituto de Química - Caixa Postal 6154 - CEP 13084-971 - Campinas - SP - Brasil.

Palavras Chave: Desenvolvimento de algoritmo, alinhamento, técnicas instrumentais

Introdução

Existe uma tendência crescente para aplicações de técnicas instrumentais visando a utilização de vetores, matrizes e tensores de dados por amostra, ao invés de um único escalar. Para extração de informações úteis destes grandes conjuntos de dados é necessária a utilização de estatística multivariada. Assim, é importante que observações decorrentes de mesmas propriedades em diferentes medidas estejam perfeitamente agrupadas. Em dados cromatográficos, picos da mesma substância, em diferentes cromatogramas, inevitavelmente devem estar "alinhados". Se o arranjo de dados não estiver disposto de forma onde as variáveis observadas das amostras submetidas à análise expressem os mesmos atributos, o pré-requisito para se utilizar modelos quimiométricos, é violado. Na literatura, observa-se que os métodos utilizados atualmente para realização de alinhamento apresentam grande consumo de tempo (algumas horas) e nem sempre resultam em alinhamento eficiente (deformações de picos). Neste trabalho é proposto um algoritmo que possibilita o alinhamento de vetores de dados (ex.: um perfil cromatográfico) e de matrizes de dados (ex.: imagens de cromatografia bidimensional). O algoritmo está baseado na maximização do coeficiente de correlação entre dados de amostras que apresentem alguma similaridade e foi testado em diferentes tipos de dados previamente obtidos por cromatografia gasosa (GC) e por cromatografia gasosa bidimensional abrangente (GCxGC).

Resultados e Discussão

A avaliação é realizada em dados de amostras desalinhadas em função de uma amostra referência (selecionada pelo analista). Para que haja uma correção o algoritmo realiza um teste de hipótese onde a amostra teste é deslocada para nos dois sentidos do vetor de dados. Se for confirmado o desalinhamento, os dados teste são corrigidos de forma interativa ao longo de vetor. Um parâmetro de incremento define onde haverá a próxima avaliação e está relacionado com a taxa de desalinhamento observada. Um segundo parâmetro chamado faixa

de análise, define o comprimento do vetor ou janela da matriz de dados que é avaliada a cada incremento e depende da similaridade das amostras. Um terceiro parâmetro define o limite superior de intensidade do sinal, em que as correções são aplicadas, de forma preservar as informações relevantes dos dados em estudo. Uma análise sobre a qualidade do alinhamento pode ser realizada baseando-se na curva de Pearson visando máxima correlação com a referência e o mínimo sobreajuste. Os resultados mostraram que o método proposto é bastante eficiente quando comparado com os programas para alinhamento disponíveis na literatura, tanto em relação ao consumo de tempo (na ordem de segundos), quanto em comparação ao alinhamento resultante. A **Figura 1** apresenta uma sobreposição de uma região de cromatogramas naturalmente desalinhados e em seguida alinhados pelo algoritmo desenvolvido.

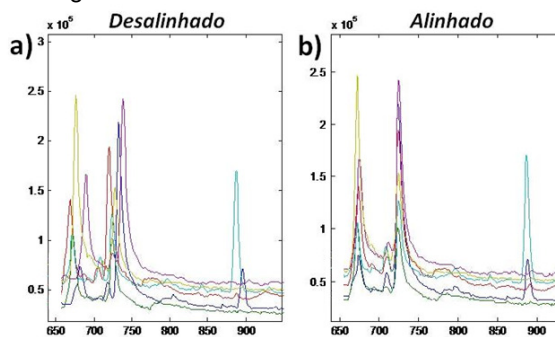


Figura 1. Cromatogramas (a) desalinhados (b) alinhados pelo algoritmo.

Conclusões

O algoritmo desenvolvido é extremamente rápido e eficaz, podendo ser utilizado também para cromatogramas bidimensionais. Esta abordagem é promissora e pode ser aplicada em diversas áreas de pesquisa, tais como as chamadas "Omics".

Agradecimentos

Os autores agradecem ao apoio financeiro da CAPES, CNPq e FAPESP.