

Abstract

Motivation

Liquid chromatography coupled to mass spectrometry (LC/MS) became a routinely used technique in differential proteomics studies. LC/MS proteomics data is processed in multiple stages. An important processing step is the retention time alignment, which aims at correcting non-linear retention time shifts between LC/MS runs [podwojski_retention_2009]. Numerous algorithms have been previously presented to solve this problem. They are usually based on the Dynamic Time Warping (DTW) and have quadratic computational complexity for memory and also computing time.

Results

We present an intuitive and efficient algorithm for pairwise retention time alignment between LC/MS peak lists in linear time and space. The algorithm combines previously introduced techniques. Using the idea of Hirschberg algorithm for the Longest Common Substring (LCS) problem, DTW's space complexity is reduced to linear [hirschberg_linear_1975]. In combination with the iterative refinement windowing techniques inspired by FastDTW [salvador_fastdtw:_2007], we also linearize the algorithm's time complexity. Similar to FastDTW we precalculate coarse time alignments on downsampled peak lists and iteratively use these prealignments to reduce the search space at the next sampling stage, eventually aligning complete peak lists. The algorithm performance can be controlled by radius parameter. The downsampling of peak lists is done by selecting representative peaks by their masses.

Discussion

The mathematical basis of Hirschberg's LCS algorithm enables parallelized computation of the alignment utilizing recent multiprocessor or GPU based techniques.