Retention time alignment algorithms for LC/MS data

Shelley Herbrich

Abstract

The goal of proteomics is to understand the individual proteins and their roles in larger biological systems. The study of the human proteome is a crucial component in fields such as drug development and biomarker discovery. Proteomics employs the liquid-chromatography mass-spectrometry technique to identify and quantify the proteins of a complex biological sample. The primary obstacle in comparative proteomics is the alignment of the LC-MS outputs with respect to retention time. In this study, linear and loess regression algorithms presented by Podwojski [2009] are constructed and implemented in Stata for a randomly generated data set. The linear regression approach is only able to accommodate linear shifts in retention time, whereas loess regression allows for the correction of non-linear shifts. In the absence of a deterministic model, a new approximation for Akaike's information criterion is introduced to determine the bandwidth of each loess regression. The data is then corrected according to both regression algorithms. The effective alignment of each is analyzed to demonstrate the advantage of algorithms that adjust non-linear shifts.