8-2016

Advancing Algorithms for Mass Spectrometry Data Analysis

Rob Smith
University of Montana, Missoula

Technical Report
UGP 2015: Advancing Algorithms for Mass Spectrometry Data Analysis

Rob Smith
Department of Computer Science
University of Montana
August 1, 2016

Abstract:
Our objective with this award was to conduct the preliminary experiments necessary to assess the state of the art of existing mass spectrometry signal segmentation algorithms. Although unforeseeable complications prevented the completion of all proposed research goals, completed research aims resulted in:

- A cross-platform mass spectrometry data viewer with the ability to pan, zoom, and scroll.
- Motivation for experimental ground truth data through exposing breakdown cases of the leading mass spectrometry simulator.
- A completed code framework for running and testing existing isotopic trace segmentation algorithms.

Since the receipt of this award, we have submitted three related external research proposals. One has been awarded (NSF CAREER award for $742,000), and one is in review.

Cross-platform mass spectrometry viewer:
Prior to this research, there was no mass spectrometry viewing software available that provided the ability to efficiently navigate through a 3-d visualization of mass spectrometry data. We developed software called JS-MS that provides a cross-platform, browser-based viewer for mass spectrometry data.

JS-MS features both a 3-d view and a 2-d view, plotting data points with colors corresponding to their intensities. The viewer is designed modularly, and works with an unspecified backend data storage module. The viewer accepts navigation requests from the user and translates these into queries to the backend. The viewer then plots the data received in response to the query. Zooming with JS-MS is initiated by click-dragging the mouse or by using the mouse scroll wheel. The zoomed display region contains more visible data points, allowing more precise analysis. All viewing features, including tracking and panning, are retained in zoomed view. The user can also use the arrow buttons to scroll up, down, left, or right in the data, a feature not available in any other mass spectrometry viewing software.

The default color palette can be toggled to a color-blind-friendly color palette using an interface button. The color blind option and other user settings are persisted in a cookie.

A legend is displayed in a corner of the graph. The legend shows the three axes with labels in the orientation matching the graph's. It can also be used to show the current viewing window in relation to the whole data set, much like a minimap does in video games. This window provides context to the local view and facilitates sequential navigation through the entire output.

A peer-reviewed short paper describing this software with two undergraduate co-authors was accepted as a podium presentation at BIOT 2015, an international bioinformatics conference.

Motivation for experimental ground truth:
Mass spectrometry simulation is complicated. We attempted to use JAMSS, a published mass spectrometry simulator, to create our ground truth data sets. After creating JS-MS, we were disappointed to find that although JAMSS has been successfully used to generate small mass spectrometry simulations with more realistic data characteristics than any other available simulator, the

signals generated in the simulations for this project were readily recognizable as unrealistic and therefore unusable for quantitative evaluation. This result was unexpected but still useful for two reasons. First, we showed that current simulation tools are insufficient for quantitative evaluation of mass spectrometry algorithms. This result informed our request for $70,000 on our NSF CAREER award for the purpose of creating ground truth data sets experimentally. Without our result, it is unclear whether we could have justified the large expense needed to generate experimental ground truth data. Second, we identified future research in the form of improving the data generation models used in JAMSS.

Framework for isotopic trace segmentation:
We created scaffolds for running parameter searches for the following published algorithms for isotopic trace segmentation: Massifquant, MatchedFilter, and CentWave. The three isotopic trace segmentation algorithms tested each require the user to provide several parameters which have a significant effect on the performance of the algorithm. In order to provide a sufficient evaluation of the performance of each method, an extensive list of parameter permutations must be tested. In other words, rather than consisting of a single run, the testing of one algorithm on one data set requires hundreds of runs, one for each permutation of possible parameters. We have developed the list of parameters and code to perform an automatic grid search of all parameter permutations will enable future quantitative evaluation of these algorithms as soon as a ground truth dataset is available. As noted, our current NSF funding provides for the construction of this dataset, and we will be able to pursue these tests shortly.