A mathematical research methodology for analysis in high throughput proteomics

Brian L. LaMarche, Stephen J. Callister, Anuj R. Shah, Aaron T. Wright, Michael J. Wilkins, Mathew E. Monroe, Kevin Crowell, Gordon A. Anderson, and Richard D. Smith
Pacific Northwest National Laboratory, Richland, WA

Introduction

The Minkowski-Bouligand (MB) dimension \( D \) is an approximation of fractal dimension. It can be used to describe the distribution of data in space.

We calculate a MB dimension for a given dataset using the box-counting algorithm. This dimension is useful to alignment because MB can provide a single value describing how a dataset covers mass and LC retention time dimensions.

To demonstrate this mathematical approach, LC-MS features were extracted from 566 LC-MS datasets of tryptically digested Shewanella oneidensis MR1 whole cell lysates. Each dataset contained 51,000 accurate mass and time tags filtered by a Peptide Prophet score of 1.08. Matches were constrained by STAC (Statistical Tools for AMT tag Confidence) scores to ensure confident matches to confident tags.

Space filling quantification

To show the robustness of MB to the number of LC-MS features, we randomly removed LC-MS features from a Shewanella dataset up to 50%. Although 50% of the features are ultimately removed, the MB is only slightly affected.

For right we show how the shape of a LC-MS feature map changes with respect to the MB dimension. MB value increases starting upper left moving down and right.

Conclusions

• The Minkowski-Bouligand metric is independent of instrumentation and thus useful as a global descriptor.

• The box-counting method converges after a few iterations; calculation of the Minkowski-Bouligand dimension scales linearly in time and space complexity.

Acknowledgements

The research was supported by a grant under the LDRD Program at the Pacific Northwest National Laboratory (PNNL). PNNL is a multiprogram national laboratory managed by Battelle for the U.S. Department of Energy under Contract DE-AC05-76RL01830.

References


CONTACT: Brian LaMarche
Environmental Molecular Science Laboratory, K-8
Pacific Northwest National Laboratory
P.O. Box 999, Richland, WA 99352
E-mail: blamarche@pnl.gov