Session 2
7th May
10:15 h

Development of computational tools for the analysis of 2D-nuclear magnetic resonance data

Bruno Pereira[1], Marcelo Maraschin[2] and Miguel Rocha[1]

[1] Centre of Biological Engineering, University of Minho, Braga, Portugal
[2] School of Agricultural Sciences, Federal University of Santa Catarina, Santa Catarina, Brazil

Metabolomics is one of the omics' sciences that has been gaining a lot of interest due to its potential on correlating an organism's biochemical activity and its phenotype. The main techniques that collect data are based on mass spectrometry and nuclear magnetic resonance (NMR) spectroscopy. The last one has the advantage of analysing a sample *in vivo* without damaging it and while its sensitivity is pointed out as a disadvantage, multidimensional NMR delivers a solution to this issue. It adds layers of information, generating new data that requires advanced bioinformatics methods to extract biological meaning. The need to establish an integrated framework has become imperative due to different approaches that multidimensional NMR has, to tackle reproducibility issues across research groups.

In recent work from the host group, *specmine*, an R package for metabolomics and spectral data analysis/mining, has been developed and improved to wrap and deliver key metabolomic methods that allow a researcher to perform a complete analysis. Tools integrated in *specmine* were developed to read, visualize, and analyse two-dimensional (2D) NMR. A new *specmine* structure was created for this type of data, easing interpretation and data visualization. In terms of visualization a novel approach towards three-dimensional environments enables users to interact with their data. The selection of which samples to plot, when the user does not specify an input, is based on a signal-to-noise ratio scale and a method to perform peak detection on 2D NMR based on local maximum search was implemented to obtain a data structure that best benefits from *specmine*'s functionalities. These include pre-processing, univariate, and multivariate analysis as well as machine learning and feature selection methods.

The 2D NMR functions were validated using experimental data from two scientific papers, available on metabolomic databases. These data originated two case studies from different NMR sources, Bruker and Varian, which reinforces *specmine*'s flexibility. The case studies were carried out using mainly *specmine* and other packages for specific processing steps. A pipeline to analyse 2D NMR was added to *specmine*, in a form of a vignette, to provide a guideline for the newly developed functionalities.