SUPERVISED CLASSIFICATION METHODS IN NMR FOR RESONANCE MAPPING OF IDPS

Javier A. Romero, Paulina Putko, Krzysztof Kazimierczuk, Anna Zawadzka-Kazimierczuk

Centre of New Technologies, Stefana Banacha 2c, Warsaw, Poland
j.romero@cent.uw.edu.pl

Machine Learning (ML) methods are being used more and more in every branch of science, and NMR is no exception. Famous examples include deep neural networks for peak picking and spectral deconvolution [1] and AUTOMAP for image reconstruction in MRI [2]. But as the field of ML keeps expanding and gaining worldwide recognition, its methods and techniques are being incorporated into every-day research tasks. I will present the case of resonance assignment in Intrinsically Disordered Proteins (IDPs). IDPs play an essential biological role, but their research is challenging. The sequential assignment procedure can be particularly difficult since the set of heteronuclear spectra used are usually crowded and signals overlap. Therefore, the chains of sequentially-linked residues are short and to map them on the protein's sequence one has to efficiently recognize which amino-acid types the residues correspond to. This is essentially a classification problem: we aim at assigning amino acid types to resonance signals (residues).

I will show how the mapping process can be assisted, to the point of almost fully automation, by basic ML classification methods. In particular, I will present Linear Discriminant Analysis (LDA) and its close relative Quadratic Discriminant Analysis (QDA), along with K-Nearest Neighbors (KNN) and Support Vector Machines (SVM). I will show how each of these methods can be used for protein mapping, how to choose a consistent training set and define criteria to choose which is the best method for classification.

REFERENCES

[1] D. W. Li, A. L. Hansen, C. Yuan, L. Bruschweiler-Li & R. Bruschweiler, *DEEP picker is a deep neural network for accurate deconvolution of complex two-dimensional NMR spectra, Nature communications* **2021**, *12*, 1–13.

[2] B. Zhu, J. Z. Liu, S. F. Cauley, B. R. Rosen & M. S. Rosen, *Image reconstruction by domain-transform manifold learning, Nature* **2018**, *555*, 487–492.