Functional data analysis and fuzzy classification. Independent concepts or a successful combination?

Itziar Irigoien[1], Concepción Arenas[2]

[1]University of the Basque Country (UPV/EHU), Department of Computation Science and Artificial Intelligence; [2]University of Barcelona (UB), Statistics Section of the Department of Genetics, Microbiology and Statistics

Nowadays we are increasingly able to collect more complex data, and many challenges in data analysis stem from that complexity. The progression from a single numerical value as the unit of study, to a multivariate vector, then to a functional curve, or even to a skeletal shape representation, illustrates this evolution. In other words, there has been a shift from using large sample sizes in low-dimensional spaces to using relatively small ones in high-dimensional spaces. The perspective offered by functional data analysis (FDA) often provides a framework that allows the analysis of curves, images, or functions in high dimensions overcoming the problem of high dimensionality. For this reason, FDA has started to appear in the computational and bioinformatics literature over the last years. On the other hand, fuzzy classification assigns a degree of membership to each unit, often used in disease diagnosis to classify patients based on medical data and with artificial intelligence techniques to address uncertainty in diagnosis. Using the COVID-19 Raman spectroscopy data set we show the usefulness of combining functional data analysis and the distance-based fuzzy classifier FC-DF highlighting their strengths and limitations.

Keywords: Functional Data, Fuzzy Classification, Depth Function, Distance-based approach.