Maximize Generalizability with Minimum Effort: Using Physical and Chemical Properties to Identify the Best Formulation Components

Mahmoud Hammoud

SAS Institute, JMP Division
mahmoud.hammoud@jmp.com

Testing different raw materials to choose the most suitable one to be included in a formulation is a typical situation that arises in various chemical manufacturing applications. A common approach consists of testing all options in an experiment, where each option is treated as its own category (material A, B, C, etc.). This requires time consuming and often expensive physical testing. Moreover, any statistical model that is built based on the resulting data is only able to predict the tested options and does not apply for "new" materials with different physical/chemical properties. An alternative and more efficient approach is to represent each raw material option by its characteristic physical/chemical properties, which is also the basic principle behind QSAR. By doing so, only a subset of the available materials have to be tested and the resulting quantitative model can be used to discover an optimal chemical composition that might not even exist yet. This optimum can be used as a benchmark when choosing among the commercially available alternatives. In an ideal scenario, one could use the result to make a case for the supplier actually producing the optimal solution. This poster illustrates an analytical workflow for applying the method described above with emphasis on dimensionality reduction techniques like Principal Components Analysis (PCA) and Partial Least Squares (PLS) which are indispensable when dealing with a large number of (correlated) predictors as is typically the case for physicochemical properties of materials. The analyses and graphics are generated using JMP, a statistical discovery software from SAS that offers a unique blend of interactive visualization and statistical analysis.