[P35] Visualization of the Photovoltaic Space by Dimension Reduction Methods

Omer Kaspi,[a] Abraham Yosipof,[b]* and Hanoch Senderowitz[a]*

[a] Department of Chemistry, Bar-Ilan University, Ramat-Gan 5290002, Israel
[b] Department of Information Systems and Department of Business Administration, College of Law & Business, Ramat-Gan , Israel

Material informatics is engaged with the application of informatics principles such as data mining and machine learning methods to analyze material science data bases and the design of new materials. A typical database is characterized by two main characteristics: the number of observations (i.e, the number of data samples) and the number of attributes (i.e, the dimensionality of the data). When dimensionality increases, data becomes more complex to analyze and increasingly sparse. In addition, density and distance between objects in the multi-dimensional space, that are critical to clustering and outlier removal, become less meaningful. Furthermore, the possible combinations of subspaces will grow exponentially making it hard for traditional analyzing methods.

In order to accelerate the data mining procedure, dimension reduction methods may come in handy, as one of the key benefits of dimensionality reduction methods is the visualization of data with high dimensionality. Dimensionality reduction methods seek to produce a 2D or 3D representation of the data, which in turn can be plotted and visualized. Their aim is to preserve as much of the significant structure of the high-dimensional data as possible in the low-dimensional map.

Here we present the first implementation and comparison of four dimension reduction methods, namely PCA, PCA kernel, Isomap and Diffusion map for the visualization and the representation of the photovoltaic (PV) space. In order to meet the research goal five previously published PV libraries were integrated to a single database. The integrated data base contains 1,165 cells (observations) and is characterized by seven experimentally measured photovoltaic properties namely J_{SC}, The short circuit photocurrent density; V_{OC} ,The open circuit photovoltage; *IQE*, The internal quantum efficiency; P_{MAX} ,The maximum photovoltaic power producible by a solar cell; *FF*, The fill factor which is the available power at the maximum power point divided by V_{OC} and J_{SC}; R_S , Series resistance and R_{sh}, Shunt resistance. The integrated database was assigned to each of the different dimension reduction methods and 3D representations were obtained. The 3D representations were evaluated by standard parameters such as the trust of the low-dimensional embedding and the 1-nearest neighbor (1NN), 3-nearest neighbors (3NN) and 5-nearest neighbor (5NN) classifiers.

The results showed that the methods captured much of the local information of the high-dimensional data very well and maintain the cluster structure of the data in the low dimension map.