[P12] Data Mining and Machine Learning Tools for Combinatorial Material Science of All-Oxide Photovoltaic Cells

Abraham Yosipof, Assaf Y. Anderson, Hannah Noa Barad, Sven Rühle, Arie Zaban and Hanoch Senderowitz

Dept. of Chemistry, Center for Nanotechnology & Advanced Materials, Bar Ilan University, 52900 Ramat Gan, Israel

Recently a new field in photovoltaics (PV) is emerging focusing on solar cells that are entirely based on metal oxide (MO) semiconductors. All-Oxide based photovoltaics comprise the potential for extremely low cost solar cells, provided they can show an order of magnitude improvement in their power conversion efficiencies. The development of such cells could benefit from combining combinatorial material sciences of compounds with data mining and machine learning tools to analyze and rationalize the empirical results and for the design of new compounds.

Here we describe what we believe to be the first reported application of data mining and machine learning techniques in the field of photovoltaics. In order to establish the tools in this new field, we focused our attention on a library of 169 solar cells[1]. These were generated on precut glass substrates onto which a compact TiO_2 window layer with a linear gradient was deposited by spray pyrolysis, followed by Pulsed Laser Deposition of Cu-O light absorber. This process led to cells with different ratios of the window and absorber layers and variable composition of the Cu-O. Each cell was characterized by 7 experimentally measured independent variables (i.e., descriptors), namely, the thickness of the TiO_2 layer, the thickness of the Cu-O layer, the thickness ratio between the two layers, the distance of the cell from the center of the Cu-O depositing plume, the band gap of the absorber, the resistance and the maximum calculated photocurrent and by 3 independent variables, namely, the short circuit photocurrent (Jsc), the open circuit photovoltage (VOC) and the internal quantum efficiency (IQE).

Following characterization, cells were divided into training and test sets using a newly developed representativeness function. Individual models were derived for all independent variables with various techniques including *k*NN and genetic programming. Typically, good models were obtained with Q^2 values for training set compounds and R^2 values for test set compounds > 0.8. Analysis of the resulting models shed light on the relative importance of the different descriptors. In the case of VOC, no models could be derived by these methods presumably due to the presence of outliers. This problem was resolved by removing ten outliers, using an in-house developed outlier removal algorithm. The remaining compounds were re-subjected to the model building procedures yielding this time reliable models. Interestingly, all outliers corresponded to library cells characterized by a thin Cu-O layer.

In summary we have demonstrated the usefulness of data mining and machine learning tools in the field of photovoltaics.

[1] A.Y. Anderson, Y. Bouhadana, H.-N. Barad, B. Kupfer, E. Rosh-Hodesh, H. Aviv, Y.R. Tischler, S. Rühle, A. Zaban, ACS Combinatorial Science 16 (2014) 53.