Materials Informatics: The marriage of materials and data sciences

Omer KASPI,[a] Hadar BINYAMIN,[a] Abraham YOSIPOF,[b] Hanoch SENDEROWITZ[a]

[a] *Department of Chemistry, Bar Ilan University, Ramat Gan, 5290002, Israel*
[b] *Department of Information Systems, College of Law & Business, Ramat-Gan, P.O.Box 852 Bnei Brak 5110801, Israel*

Materials informatics is rapidly developing as evidenced by the large increase in the number of publications in this field. This development is fueled by the continuous growth of available data, both experimental and computational, on one hand and by the ability to draw from the rich repertoire of methods in its sister field, chemoinformatics, on the other. Similar to chemoinformatics, materials informatics can help bridge the so-called data-knowledge gap by offering smart ways to navigate the enormous materials space in search of new materials with favorable properties.

A key computational technique in materials informatics is Machine Learning (ML). In this talk, we will therefore comparatively analyze the similarities and differences between materials informatics and cheminformatics, within the framework of what is required to derive reliable ML-based models.

Next we will focus on the application of ML methods to the study of solar cells. Such cells hold the potential to meet the growing worldwide demand for clean energy. Today most solar cells are based on silicon yet new alternatives are continuously emerging including organic photovoltaic cells, dye sensitized solar cells (DSSCs), and metal-oxide (MO)-based solar cells. However despite significant progress, all these cells could benefit from improvements in key components, e.g., the dye in DSSCs and the MO composition of MO-based solar cells. Thus, we will present how the concept of pharmacophore can help identify new dyes with favorable (predicted) electronic properties for DSSCs and how combining combinatorial materials sciences with ML can lead to predictive models for key solar cells parameters such as current, voltage and quantum efficiency.

Finally, we will shift our attention to the field of forensic informatics and in particular to the usage of ML in order to analyze physical evidence found in crime scenes. This work will highlight the usefulness of experimentally determined spectra, in particular such that reflect the elemental composition of such evidence. In particular, we will describe the development of a reliable of ML-based classifier from glass fragments retrieved from different types of car windshields using ion beam analysis.

As a final take home message, we emphasize the need to conduct research in the field of materials informatics in close collaboration with experimentalists in order to provide insight into the observed trends and to capitalize on the results.