Data Driven Design of Perovskites

Hongbin Liu[1], Yongquan Xie[2], Sarah D. Floris[1]

[1] Department of Chemistry, University of Washington, Seattle, WA 98195

[2] Department of Material Science & Engineering, University of Washington, Seattle, WA 98195

Objective

We are working on finding the optimal formula for the hybrid organic inorganic perovskites (HOIPs) for solar cell energy conversion by using statistical models to analyze scattered data in published articles.

Hybrid Organic Inorganic Perovskites

In recent years, there has seen a rapid emergence of new class of solar cell based on HOIPs. Incorporating perovskites into semiconductor devices such as solar cells has shown good performance. Although the first efficient solid-state perovskite cells were reported in 2012, a very fast progress was made during five years with power conversion efficiencies (PCE) reaching a confirmed 20%. [1]

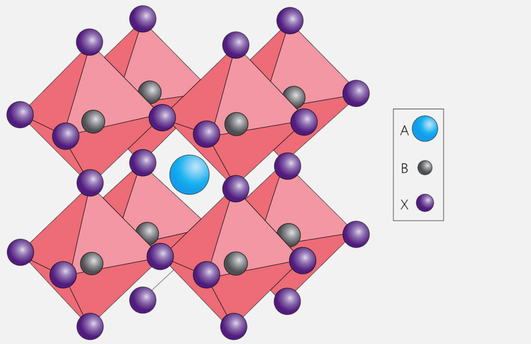

Figure 1: The schematic structure of a ABX_3 HOIP. Usually A sites are cations, B site is Pb^{2+}, and X sites are halogens.

Workflow

Despite having many studies on mixtures of HOIPs, data is not recorded in a centralized database. We searched related papers from 2015 to 2017, finding 250 different formulas in about 50 papers.

Figure 2: The workflow of the project.

Optimal Components in the ABX_3 Formula

Because this project had limited amount of data, it required us to find unique ways to run efficient analysis. We compressed our origin data structure since the sparsity muddled the correlation of the data.

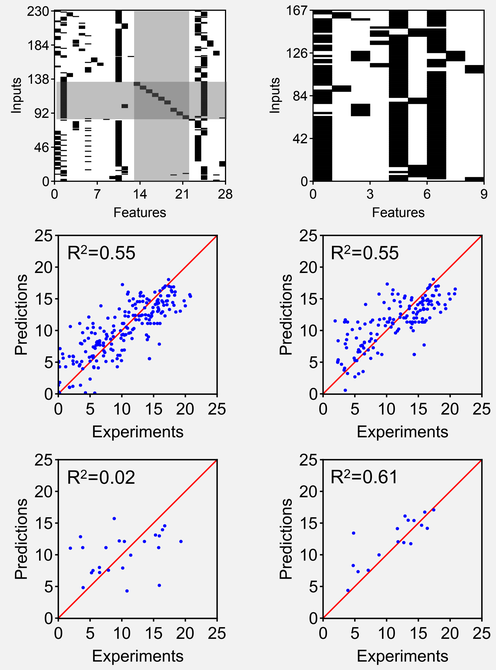

Figure 3: The parity plots based on two type of data structures.

The merit of the linear regression model (LASSO) is to give the inference of each feature to the max PCE. The positive coefficients indicate the possible components can improve the max PCE. The FA/MA, Pb and I/Br have the highest coefficients among their categories.

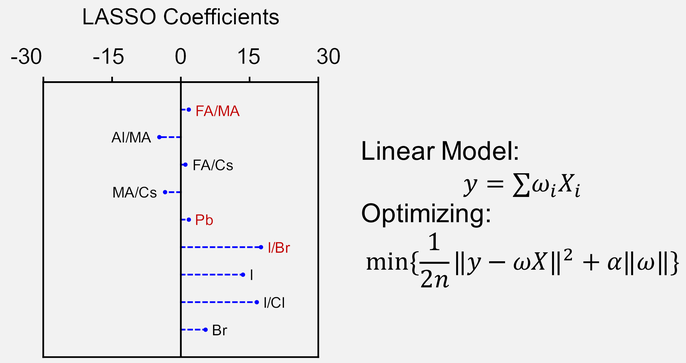

Figure 4: The fitting parameters from the LASSO.

Optimal Ratios of FA/MA and I/Br

Once the optimal components have been determined, our main focus switches to the surface fitting of the FA/MA and I/Br ratios. We have tried several non-linear regression models along with the GridSearchCV and RandomSearchCV from SciKit for determining their parameters.The back-propagated neural network turned to have the best performance among all.

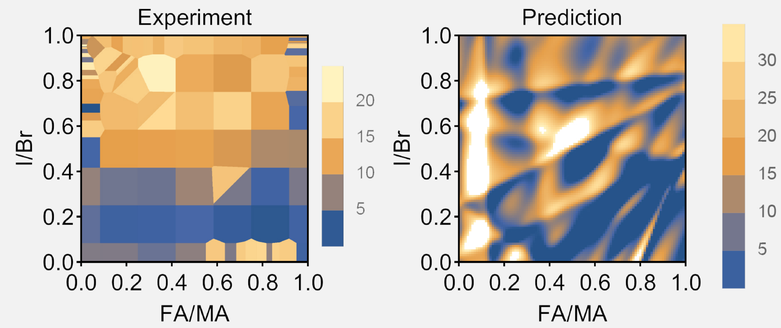

Figure 5: The prediction based on the BP-NN

Though the best fitting quality is 0.58 due to the surface curvature complexcity and the limited amount of data, a qualitatively correct trend of PCE in $FAPbI_3$, $FAPbBr_3$, $MAPbI_3$ and $MAPbBr_3$ can be observed.

Conclusion

We have used the data science techniques to identify the optimal combination of elements in the HOIPs for the energy conversion in the solar cells. The lead based HOIP is so far still the best candidate for the solar cell application. A prediction in the full range of FA/MA, I/Br ratios has been made based on the neural networks.

References

[1] W. S. Park et al, *Science* **348**, 2015, 1234-1237.

Acknowledgments

This project was supported by the Data Intensive Research Enabling Clean Technology (DIRECT) program at the University of Washington, Seattle, WA. The authors thank Prof. Xiaosong Li, Prof. David Beck, Prof. Jim Pfaendtner and Prof. Aleksandr Aravkin for the insightful discussions.