CL325: Chemical Reaction Engineering II
Course Project

Awasthi Gaurav
200020036

Materials genes of heterogeneous catalysis from clean experiments and artifcial intelligence[1]

Lucas Foppa,* Luca M. Ghiringhelli, Frank Girgsdies, Maike Hashagen, Pierre Kube, Michael Hävecker, Spencer J. Carey, Andrey Tarasov, Peter Kraus, Frank Rosowski, Robert Schlögl, Annette Trunschke, and Matthias Scheffer

MRS Bulletin 46, 1016–1026 (2021)

I. Area and nature of work

The broad area of work is heterogeneous catalysis. The problem addressed is that of modelling the complex catalytic process, identifying the key parameters determining the performance and using the analysis to guide catalyst discovery and design.
The work is experimental and computational, with the artificial intelligence model being built to find correlations in the experimental data.

II. Significance of the problem addressed

The performance of heterogeneous catalysis depends on a complex interplay of many processes – an introductory course on chemical reaction engineering giving an overview of various factors such as mass transfer limitations, adsorption/desorption mechanisms and reaction kinetics was enough to convince us of this.
Currently available experimental techniques do not help determine a general relationship between the catalyst's physicochemical properties, the reaction conditions, and its performance. Addressing this gap can help direct the search for optimised catalysts with desired properties.

III. Synopsis of the work

In this study, the reaction of interest is the selective oxidation of alkanes to olefins and oxygenates. The aim of the catalyst is to selectively form the oxygenate (acrylic acid) by partial oxidation of propane (equations 1, 2 and 3 of the article[1]).

The authors synthesise and characterise nine vanadium-based catalysts (shown in Figure 1(a) of the article[1]) to find over 40 properties per material.

They then construct the “descriptors”, which are essentially analytical expressions for the target properties. These expressions are constructed by considering various mathematical operations on the input parameters (addition, absolute difference, multiplication, logarithm etc.). This generates billions of possible descriptor candidates (since the space of descriptor candidates grows combinatorially with the number of primary features and the multiple mathematical operations). The properties calculated from these analytic expressions are then compared with what was experimentally determined, and an elementary RMSE minimisation provides the most accurate descriptor. In this study, the target property was chosen as the selectivity of the catalyst towards the formation of acrylic acid.

The final descriptor is shown in equation 5 of the article[1].
Inspecting this optimal descriptor also provides insights into the “materials genes” (the key parameters that determine catalyst performance). The catalyst genes identified include those related to porous structure (pointing to mass transfer/diffusion-related processes), the electric potential on the surface of the catalyst, the concentration of vanadium at the surface and the amount and type of surface carbon.
A similar analysis was performed with the target property being the conversion of propane. The descriptor for this property pointed to some different catalyst genes, as expected. Some genes, however, are common, indicating that selectivity and conversion are not wholly unrelated and that there are some base catalyst properties which influence all processes.

An interesting application of this study is the generation of “maps of catalysts”. These describe the target property as a function of the descriptor and the temperature. Not only does it summarise the relative performances of the catalysts used in the study, but it can also be used to guide the design and discovery of new catalysts since we now understand the properties most important to our end goal.

IV. Critique of work

I personally found the authors’ method of training and validating the model very elegant. They provide an excellent workaround to the scarce training data for the AI model – while there are only nine catalysts, the large number of properties and descriptors capture the underlying catalyst dynamics adequately. To validate this model, they used a method called “leave-one-material-out cross-validation”. The model is trained with the dataset corresponding to eight of the catalysts, and the resulting equations are tested upon the ninth catalyst. This is repeated such that each catalyst is left out once. The RMSEs corresponding to the models are averaged over all iterations, and the optimal model is the one with the minimum RMSE. I found this an interesting analogue to the conventional train-test-validate method adopted in machine learning problems, where the dataset is split into three parts for training, testing and validating the model.

One query, whose explanation I could not find in the paper or in the references was how the authors account for dimensional validity while constructing the descriptors. For instance, any descriptors which involve the addition of say, mole fraction and temperature, should immediately become invalid. Removing all such invalid descriptors could substantially decrease the available descriptor space for training and validating in the model.

V. Writing style

The paper is well-written. A particular stand-out quality for me was how the authors provide a brief description regarding each novel term or process in the text itself, while of course referencing more detailed resources. This helps a novice reader maintain continuity while reading without having to redirect themselves to several resources to understand basic terminology.

VI. Final Conclusion

Credit is due to the authors for their detailed analysis and succinct explanations. I was very interested in reading the code behind the AI model, which they have linked in the paper; unfortunately, that portion of the website does not appear to be working currently[2]. The methods described in this paper are part of a much larger goal to guide the discovery of new materials[3]. The cumulative resource built seems impressive, allowing users to query for material properties and providing AI toolkits along with explanatory tutorials.
Despite having no formal exposure to this field, apart from some elementary knowledge of catalysis through a course on chemical reaction engineering, I found the paper a fascinating read.

VII. References

1. Foppa, L., Ghiringhelli, L.M., Girgsdies, F. et al. Materials genes of heterogeneous catalysis from clean experiments and artificial intelligence. MRS Bulletin 46, 1016–1026 (2021).
2. https://nomad-lab.eu/AItutorials/PropaneOxidation
3. https://cms.nomad-lab.eu/about/scope