Machine learning analysis of photocatalytic CO_2 reduction on perovskite materials

İrem Gülçin Zırhlıoğlu , Ramazan Yıldırım
Bogazici University, Department of Chemical Engineering, Istanbul, 34342, Turkey

In recent years, the CO_2 emission to atmosphere has been increased with the increasing of the fossil fuel consumption and industrial developments. For a sustainable world, precautions such as reducing fossil fuels consumption, use of renewable energy resources and converting CO_2 to other more valuable chemicals should be considered (Kumar et al., 2020).

In this study, the aim was to analyze the photocatalytic CO_2 reduction on perovskites using machine learning on the dataset constructing from the literature until March 2023. The dataset consisted of 309 data from 62 articles; the samples were examined in the light of 29 variables (descriptors) and the total production yield ($\mu mol.g^{-1}.h^{-1}$) was used as the target variable.

The bandgap was predicted via linear regression and the yield analysis was done via decision tree, random forest, gradient descent, and association rule mining.

Linear regression was used to predict the missing bandgap values on python. The rest of the machine learning applications were done in R language. The decision tree was used to classify the total yield by *rpart* and *rattle* algorithms. The random forest was used to predict total yield by regression and *randomForest* used as the library. The gradient boosting library, *xgboost*, was also used to predict the total yield by regression. The association rule mining was done by the library, *arules* and the relation between the features and the output was obtained.

The dataset was divided into 2 parts according to the reaction phase (gas and liquid) and analysis was done separately.

For the bandgap prediction by linear regression, R-square are 0.97 and 0.75 for train and validation sets respectively as in Figure 1. The fitness of the model was quite good.

Association rule mining analysis shows that some properties of the experiments such as sonication as the synthesis method, short calcination time, Ru as the cocatalyst or Ba as a doped material have high probability to have high yield.

In the decision tree for the total yield classification, the cp and the minsplit was chosen as 0.015 and 5 for gas and liquid phase datasets. For gas data, the accuracy for train and test set are 0.89 and 0.76 respectively while these values are 0.87 and 0.84 for liquid data. The conditions leading to low total yield were are also predictable for both phase. The feature importance analysis shows that perovskite synthesis method and the type of perovskite are the most important features for gas and liquid phase respectively.

For random forest, R-square values are 0.77 and 0.64 for train and test set in gas phase respectively while these values are 0.77 and 0.49 in liquid phase. Band gap is the most effective feature for gas phase predictions while this was the cocatalyst in liquid phase.

Gradient boosting shows higher performance. For gas phase, the R^2 for training and testing were 0.95 and 0.65 respectively while they were 0.94 and 0.79and for liquid phase.

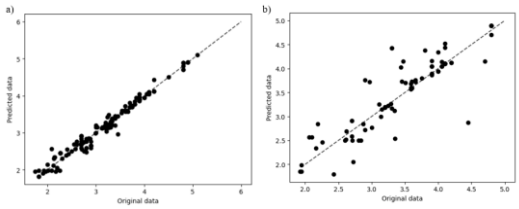

Figure 1. The original data versus predicted data for a) train set b) validation set

References

Kumar, A., Kumar, A., & Krishnan, V. (2020). Perovskite Oxide Based Materials for Energy and Environment-Oriented Photocatalysis. *ACS Catalysis*, *10*(17), 10253–10315.

İrem Gülçin Zırhlıoğlu is a master student in the Department of Chemical Engineering at Boğaziçi University. She obtained her bachelor's from Boğaziçi University and started her master in 2021. Her work includes machine learning analysis of energy systems, solar energy applications, catalytic applications, etc.

İrem Gülçin Zırhlıoğlu, e-mail: irem.zirhlioglu@gmal.com tel: +90(505) 836-4902