Approaches to Construction of Chemical Reaction Image Diagnosis System Using Machine Learning

Yuki Ide
Specially Appointed Assistant Professor
Affiliation: Institute for Chemical Reaction Design and Discovery(ICReDD), Hokkaido University
e-mail: ide.yuki@icredd.hokudai.ac.jp
Specialty: Structural Organic Chemistry and Material Chemistry

ABSTRACT

Appearance information such as color and shape are usually used to distinguish compounds. When confirming the progress of chemical reactions, there are many opportunities to judge the appearance of compounds. However, precision in judging the reaction yield varies significantly depending on experimental skills and knowledge level. By making machine learning models[1] using well-defined images of compounds, we expected that anyone can easily perform the diagnosis of reaction yields with high reproducibility, instead of using unclear parameters based on empirical results.

In this presentation, I would like to talk about the construction of an imaging diagnosis system using machine learning models, collaborating with Prof. Takikawa's group (ICReDD). In order to conduct image diagnosis of reaction yields, it is necessary to discriminate quantitatively compound ratios before and after the reaction. The effective collection of images in appropriate formats for machine learning is also an important factor. We proposed to use commercially available sugar (sucrose) and table salt (NaCl: sodium chloride) to create a large number of datasets with desired mixing ratios (wt%) (**Fig. 1a**). To ensure higher diagnostic accuracy, we attempted to normalize the imaging capture settings. After optimization of the amount and shape of the mixture samples, image diagnosis systems embedded with the pre-trained model (using 500 images) were constructed. A third person uploads an image prepared with any mixing ratio to the diagnostic imaging system, and the predicted values are displayed on the system within a few minutes (**Fig. 1b**). It is also available to provide observed-predicted plots, MAE, RMSE, and R2 values as evaluation indicators (**Fig. 1c**). Furthermore, we can predict with high accuracy the mixture sample ratio of α-glycine and γ-glycine, which have different crystal polymorphs[2].

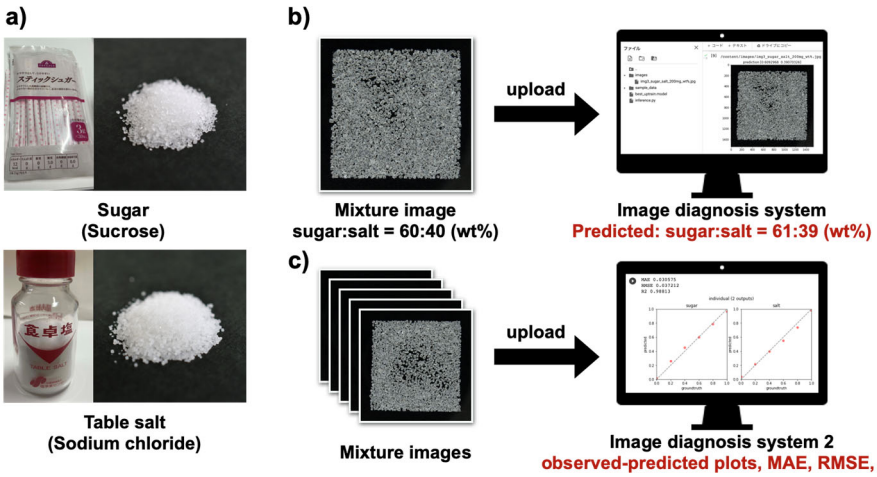

Fig. 1 Quantitative image diagnostic system using sugar and salt. **(a)** Commercial sugar and salt. **(b)** Simple image diagnosis system. **(c)** Image diagnosis system capable of exporting accuracy evaluation indicators.

1. Y. LeCun, Y. Bengio, G. Hinton, *Nature,* **2015**, *521*, 436–444.
2. G. Han, S. Thirunahari, P. S. Chow, R. B. H. Tan, *CrystEngComm,* **2013**, *15*, 1218–1224.