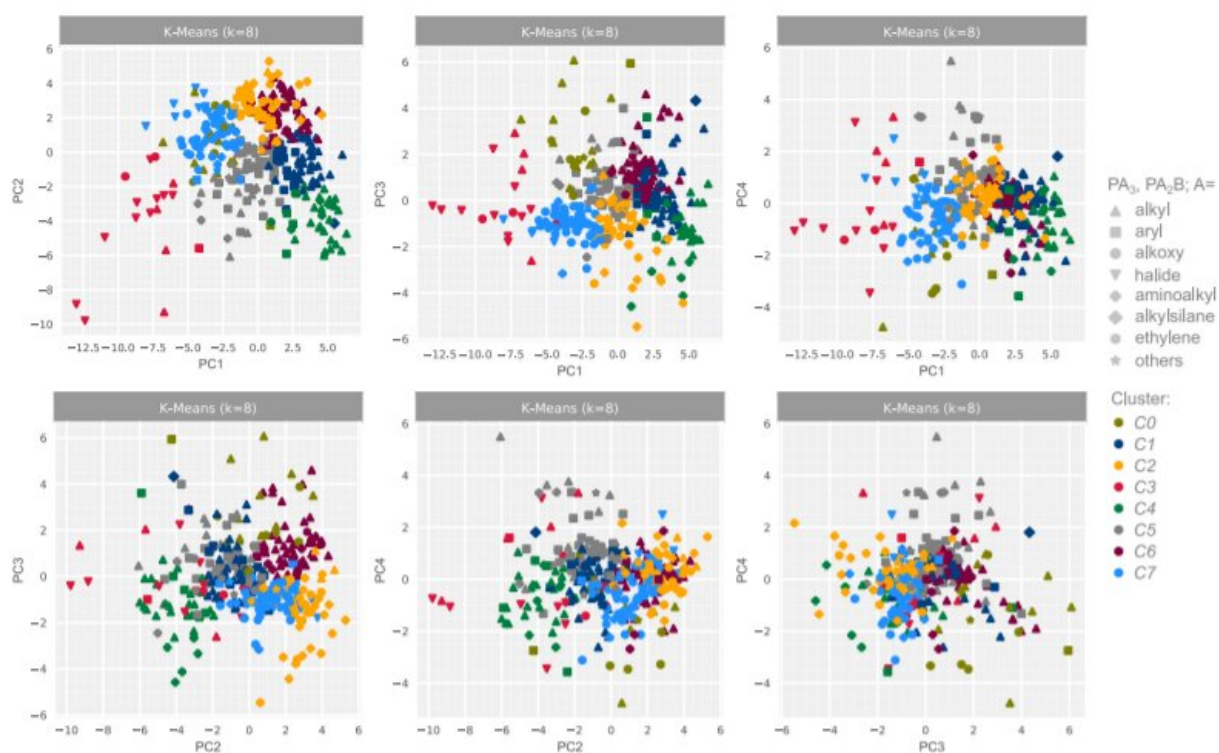# Using machine learning and computationally derived descriptors to find special classes of catalysts

December 3 2021, by Bob Yirka

Results of the initial k-means clustering (k = 8) of the LKB-P visualized using PC1 to PC4 which were adapted from reference (12). Credit: DOI: 10.1126/science.abj0999

A team of researchers at RWTH Aachen University and the University

of Jyväskylä has developed a system based on machine learning and computationally derived descriptors that can be used to find special kinds of catalysts. In their paper published in the journal *Science*, the group describes using machine learning algorithms to find patterns in known types of ligands and applying the results to find new catalysts.

Finding new catalysts that are useful to chemists is a challenging prospect. Historically, it has been done through trial and error. Over the past several years, scientists have been looking for ways to speed up the process in the hope that new catalysts can be discovered that could be used to create new and perhaps exotic products. In this new effort, the researchers have turned to machine learning to help with new searches.

The researchers trained their algorithm with examples of the general properties of known ligands. Then they used it to filter 348 ligands and sorted them into clusters using computationally derived descriptors. This resulted in a large dataset that was grouped into smaller units that could each be used for specific purposes. They then verified that their processing had worked as hoped by predicting ligands that had been synthesized before. Pleased with their results, they then used it to find special classes of catalysts. They found that they were able to create new palladium (I) dimers using catalysts.

The researchers note that the new system was able to discover new ligands using only five data points—other systems, they point out, require far more data. They note also that the reason their system is able to use so much less data than others is due to combining their database with high-quality descriptors and their two-stage clustering approach. They claim that the accuracy of their system is "remarkable"—it has already been making ligand suggestions that would likely never have been found otherwise. They further suggest that their system will likely be used to solve other problems related to finding catalysts in general.

Provided by Science X Network