Machine learning cracks the oxidation states of crystal structures

July 6 2021

Credit: CC0 Public Domain

Chemical elements make up pretty much everything in the physical world. As of 2016, we know of 118 elements, all of which can be found categorized in the famous periodic table that hangs in every chemistry

lab and classroom.

Each element in the [periodic table](#) appears as a one-, two-letter abbreviation (e.g. O for oxygen, Al for aluminum) along with its [atomic number](#), which shows how many protons there are in the element's nucleus. The number of protons is enormously important, as it also determines how many electrons orbit the nucleus, which essentially makes the element what it is and gives it its chemical properties. In short, the atomic number is an element's ID card.

The periodic table should include oxidation states

Publishing in *Nature Chemistry*, chemical engineers at EPFL's School of Basic Sciences investigate another number that must be reported for each element in the periodic table: the element's oxidation state, also known as oxidation number. Simply put, the oxidation state describes how many electrons an atom must gain or lose in order to form a [chemical bond](#) with another atom.

"In [chemistry](#), the oxidation state is always reported in the chemical name of a compound," says Professor Berend Smit who led the research. "Oxidation states play such an important role in the fundamentals of chemistry that some have argued that they should be represented as the third dimension of the periodic table." A good example is chromium: in oxidation state III it is essential to the human body; in oxidation state VI, it is extremely toxic.

Complex materials complicate things

But although figuring out the oxidation state of a single element is pretty straightforward, when it comes to compounds made up of multiple elements, things become complicated. "For complex materials, it is in

practice impossible to predict the oxidation state from first principles," says Smit. "In fact, most quantum programs require the oxidation state of the metal as input."

The current state-of-the-art in predicting oxidation states is still based on a something called "bond valence theory" developed in the early 20th century, which estimates the oxidation state of a compound based on the distances between the atoms of its constituent elements. But this doesn't always work, especially in materials with crystal structures. "It is well known that it is not only the distance that matters but also the geometry of a metal complex," says Smit. "But attempts to take this into account have not been very successful."

A machine-learning solution

Until now, that is. In the study, the researchers were able to train a machine-learning algorithm to categorize a famous group of materials, the metal-organic frameworks, by oxidation state.

The team used the Cambridge structural database, a repository of crystal structures in which the oxidation state in given in the name of the materials. "The database is very messy, with many errors and a mixture of experiments, expert guesses, and different variations of the bond valence theory are used to assign oxidation states," says Smit. "We assume that chemistry is self-correcting," he adds. "So while there are many errors on individual accounts, the community as a whole will get it right."

"We basically made a machine-learning model that has captured the collective knowledge of the chemistry community," says Kevin Jablonka, a Ph.D. student in Smit's group at EPFL. "Our machine learning is nothing more than the television game "Who Wants To Be A Millionaire?" If a chemist does not know the oxidation state, one of the

lifelines is to ask the audience of chemistry what they think the oxidation state should be. By uploading a crystal structure and our machine-learned model is the audience of chemists that will tell them what the most likely oxidation state is."

More information: Kevin Maik Jablonka et al, Using collective knowledge to assign oxidation states of metal cations in metal–organic frameworks, *Nature Chemistry* (2021). DOI: 10.1038/s41557-021-00717-y

Provided by Ecole Polytechnique Federale de Lausanne