

Machine Learning for Predicting New Chemical Compounds

Maksim Molokeev, a scientist at SibFU's School of Engineering Physics and Radio Electronics, proposed using machine learning to detect the relationship between the chemical formula of inorganic red and infrared phosphors and the wavelength of radiation, as well as the half-width of the maximum radiation.



The training was conducted on a random sample (70%) of 300 compounds. For verification, a tenfold cross-validation was used on test samples, which showed a prediction accuracy of ± 30 nm. It is a decent result since the range of predicted wavelengths is from 620 to 1030 nm. According to some experts' forecasts, the development can be used to solve a number of materials science problems, first of all, to search for chemical compounds that will be customized for specific industries and application methods.

For several years, SibFU scientists have been searching for special chemical compounds — phosphors. They are used for the production of fluorescent lamps, advertising lighting, but most importantly — for growing plants in greenhouses, including in hydroponic conditions, without losing their useful nutritional properties. Maksim Molokeev, assistant professor of the Solid State Physics and Nanotechnology Specialized Department, SibFU's School of Engineering Physics and Radio Electronics, proposed a method for selecting suitable phosphors using machine learning.

"We have systematized information about the composition and some characteristics of phosphors based on 300 scientific articles. This information array was used to train the model in a machine learning program I wrote. The program turned out to be a gifted student — now it can predict the luminescent properties of compounds using a chemical formula. It is enough to enter compositions into the program (even those that do not yet exist as chemists have not yet synthesized them), and it will predict what luminescent properties these potential compounds will have. And if these properties turn out to be promising, then it's worth doing research and making such substance in reality," said **Maksim Molokeev**.



This program can also be used to correct already known chemical compounds with good quantum yields but not with an optimal wavelength: sort of adjust the wavelength of radiation and make it more convenient for plant growth. The program will simulate a slight change in the chemical composition of the compound so that the wavelength would change as desired.

According to the developer, today many of his Russian and Chinese fellow chemists are trying to implement the forecasts of the new program and synthesize previously non-existent compounds for which the program predicted the wavelength. If they coincide within ± 30 nm, this will prove the high practical value of the newcomer. If the accuracy is lower than expected, the model will receive a new task and will be further trained.

"Of course, in a test mode, the program has already been verified with hidden data from 90 compounds, and those were exactly the compounds for which the forecast error was ± 30 nm on average, which is not bad since the wavelength in general can vary from 620 to 1030 nm.

*In fact, we took 300 compounds with known compositions and desired properties, and selected 210 of them randomly for training, having hidden the remaining 90 from the program. The program was trained on 210 samples, just like university students get their knowledge while studying. But the real assimilation of information is checked only by homework. So we set problems on prediction of the characteristics of the remaining 90 compounds but hide the correct answers. All we had to do was to check the forecast with the true values. And they matched up well!" summed up **Mr. Molokeev**.*

The scientist noted that it took only 4 days to collect the data, which opens up enormous prospects for using article data in the future. Moreover, building a model required only the chemical formula of the compounds without their complex structural parameters, which is very important since chemists can only make a reality of certain chemical formula but not sets of structural parameters such as bond lengths, bond angles, and so on. But even a student who has completed a machine learning course can use this new tool or create some of their own for forecasting.

[Пресс-служба СФУ](#), 31 january 2023

© Siberian federal university. Website editorial staff: +7 (391) 246-98-60, info@sfu-kras.ru.

Web page address: <https://news.sfu-kras.ru/node/27269>