

Учёный из СФУ создал программу-предсказатель новых химических соединений

Учёный Сибирского федерального университета Максим Молокеев предложил использовать машинное обучение для обнаружения взаимосвязи между химической формулой неорганических красных и инфракрасных люминофоров и длиной волны излучения, а также полушириной максимума излучения.



Обучение проводилось на случайной выборке (70%) из 300 соединений. Для проверки использовалась десятикратная кросс-валидация на тестовых образцах, которая показала точность прогноза ± 30 нм, что является достойным результатом, поскольку диапазон прогнозируемых длин волн составляет от 620 до 1030 нм. По прогнозам специалистов, разработку можно использовать для решения ряда материаловедческих задач, в первую очередь, для поиска химических соединений, которые будут максимально «заточены» под конкретные отрасли и способы их применения.

В СФУ уже несколько лет ведётся важная работа по поиску особых химических соединений — люминофоров. Их используют, чтобы делать лампы дневного света, рекламную подсветку, а главное — чтобы выращивать различные сельскохозяйственные растения в теплицах, в том числе, в условиях гидропоники, не теряя полезных пищевых свойств. Учёный ИИФиРЭ Максим Молокеев предложил способ подбора подходящих люминофоров с помощью машинного обучения.

«На основе 300 научных статей из различных источников была систематизирована информация о составе и некоторых характеристиках люминофоров. Этот „информационный массив“ был использован для обучения модели в написанной мной программе машинного обучения. Программа оказалась способной ученицей — теперь она умеет по химической формуле прогнозировать люминесцентные свойства соединений. При этом достаточно ввести в программу составы, даже те, которые пока не существуют — их ещё химики не синтезировали, и она спрогнозирует, какие у этих потенциальных соединений будут люминесцентные свойства. И если они интересные, то стоит заняться исследованием и такое соединение создать в реальности», — рассказал доцент базовой кафедры физики твердого тела и нанотехнологий ИИФиРЭ СФУ **Максим Молокеев**.



Такую программу можно также использовать для коррекции уже известных химических соединений с хорошими квантовыми выходами, но не оптимальной длиной волны. Как бы «подстроить» длину волны излучения и сделать её более удобной для роста растения. Программа смоделирует небольшое изменение химического состава соединения, чтобы длина волны изменилась в нужную сторону.

По словам разработчика, сейчас многие его коллеги химики в России, включая коллег из Китая, стараются воплотить «прогнозы» новой программы и синтезировать ранее не существовавшие соединения, для которых программа предсказала длину волны. Если они совпадут в пределах ± 30 нм, это докажет высокую практическую ценность новинки. Если же точность будет ниже ожидаемой, то модель получит новое задание и будет обучаться, восполняя пробелы в своих познаниях.

*«Для тестов, конечно, программа уже прошла проверку скрытыми для неё данными из 90 соединений, вот как раз для них ошибка прогноза в среднем и составляла +/- 30 нм, что неплохо, поскольку длина волны в целом может быть от 620 до 1030 нм. Фактически, взято 300 соединений с известными составами и нужными свойствами и из них выбрано случайно 210 для обучения, а оставшиеся 90 спрятаны от программы. Программа обучилась на 210 образцах, примерно как студент в университете усваивает знания на лекции. Но реальное усваивание информации проверяется только „домашним заданием“. Задачи из прогноза характеристик 90 соединений и были заданы, при этом правильные ответы были скрыты. Оставалось только сверить прогноз с истинными значениями. И они неплохо совпали!», — резюмировал **Максим Молокеев**.*

Учёный отметил, что на сбор данных понадобилось всего 4 дня, что открывает колоссальные перспективы использования статейных данных в будущем. При этом для построения модели потребовалась всего лишь химическая формула соединений и никаких сложных их структурных характеристик, что очень важно, поскольку химики могут воплотить в жизнь лишь определенную химическую формулу, а не наборы структурных параметров, таких как длина связей, валентные углы и т.д. А вот воспользоваться новым инструментом или создать такой инструмент для «предсказаний» может даже студент, прошедший курс машинного обучения.

[Пресс-служба СФУ](#), 19 января 2023 г.

© Сибирский федеральный университет. Редакция сайта: +7 (391) 246-98-60, info@sfu-kras.ru.

Адрес страницы: <https://news.sfu-kras.ru/node/27231>