

Студент НГУ впервые разработал модель глубокого машинного обучения для автоматической обработки спектров рентгеновской фотоэлектронной спектроскопии

Разработка Артема Вахрушева позволит исследователям многократно сэкономить время при обработке данных. Сейчас он работает над графическим интерфейсом этого сервиса.

Новосибирск, 02 декабря 2024 года: Нейронную сеть для автоматической обработки данных, полученных методом рентгеновской фотоэлектронной спектроскопии, разрабатывает студент 4 курса [Факультета естественных наук Новосибирского государственного университета](#) Артем Вахрушев. Также молодой исследователь работает в Институте интеллектуальной робототехники НГУ и в НТК Поверхность Института катализа им. Г.К. Борескова СО РАН. Его проект вошел в число победителей молодежного конкурса научно-исследовательских проектов «Рентгеновские, синхротронные, нейtronные методы междисциплинарных исследований». Данную работу Артем Вахрушев выполнял при финансовой поддержке Министерства науки и высшего образования РФ в рамках государственного задания ИК СО РАН (проект FWUR-2024-0032), а также при поддержке Программы «Приоритет—2030».

Рентгеновская фотоэлектронная спектроскопия (РФЭС) — полуколичественный спектроскопический метод исследования элементного состава, химического и электронного состояния атомов на поверхности изучаемого материала, основанный на явлении внешнего фотоэффекта. Спектры РФЭС получают путем облучения материала пучком рентгеновских лучей с регистрацией зависимости количества испускаемых электронов от их энергии связи.

Метод рентгеновской фотоэлектронной спектроскопии (РФЭС) — один из самых распространенных спектроскопических методов в катализе, материаловедении, физике полупроводников и многих других областях современной науки. В настоящее время обработка рентгеновских фотоэлектронных спектров в подавляющем большинстве случаев осуществляется вручную и представляет собой рутинный процесс — на описание одного спектра у научного сотрудника уходит порядка 10 минут. Классические алгоритмы анализа спектров могут вызывать значительную ошибку при обнаружении и подгонке пиков, и поэтому не могут использоваться в задачах автоматизации.

— С открытием ЦКП «Сибирский кольцевой источник фотонов» объем данных значительно увеличится, а инструментов для их автоматической обработки до сих пор нет. И мы решили впервые доверить решение этой проблемы нейросетям. Задачи обработки спектров могут быть решены методами глубокого машинного обучения, так как модели нейронных сетей способны обрабатывать такие неструктурированные данные, как изображения и последовательности сигналов, находя в них закономерности и оптимизируя ручной труд. Мы рассматривали проблему обнаружения пиковых площадей и максимумов как проблему сегментации. Для обучения модели мы использовали синтетические данные. Каждый такой синтетический спектр

включал в себя шум, пики и фон неупругого рассеяния. Пики моделировались с помощью функций псевдо-Фойгта со случайными параметрами. Фон неупругого рассеяния моделировался с помощью нормальной кумулятивной функции распределения. Затем мы разработали алгоритм для постобработки результатов анализа модели, — рассказал Артем Вахрушев.

В качестве данных для проверки модели и алгоритмов постобработки использовались спектры хлорида серебра (AgCl), полученные научной группой в Институте катализа им. Г.К. Борескова СО РАН. В результате исследований выяснилось, что разработанный Артемом Вахрушевым подход, сочетающий в себе сегментацию спектров нейросетью и алгоритмов постобработки, хорошо согласуется с результатами ручного анализа.

Разработка архитектуры модели, обучение и тестирование были выполнены с использованием PyTorch на Python. А поскольку архитектура была оптимизирована для быстрого вывода, процесс обработки спектра занимает менее секунды на домашнем ноутбуке.

В настоящее время данная нейросетевая модель существует только в виде кода в открытом доступе и неудобна для потенциальных пользователей, но Артем Вахрушев уже приступил к разработке ее графического интерфейса. Также он планирует создать потоковую обработку большого количества спектров, чтобы данная программа могла выводить динамику изменения состава исследуемой поверхности.