

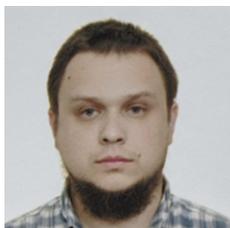
УДК 004.822-501

АВТОМАТИЗАЦИЯ ОБРАБОТКИ РЕЗУЛЬТАТОВ ИССЛЕДОВАНИЯ СТРУКТУРЫ И СВОЙСТВ НАНОМАТЕРИАЛОВ



Н.А. Шиманский

Студент кафедры физики
твёрдого тела
и нанотехнологий физического
факультета БГУ, ведущий
инженер-программист
компании
IDA Technologies (ПВТ)
nikita.shymanski@gmail.com



А.В. Баглов

Научный сотрудник НИЛ
энергоэффективных
материалов и технологий,
старший преподаватель
кафедры физики твёрдого
тела и нанотехнологий
физического факультета БГУ
baglov@bsu.by



Л.С. Хорошко

Ведущий научный сотрудник
НИЛ энергоэффективных
материалов и технологий, доцент
кафедры физики твёрдого тела и
нанотехнологий физического
факультета БГУ,
канд. физ.-мат. наук, доцент
khoroshko@bsu.by

Н.А. Шиманский

Студент 4 курса кафедры физики твёрдого тела и нанотехнологий физического факультета БГУ специальности «Физика (производственная деятельность)». Занимается проектированием и разработкой IT-решений в области бизнеса и науки. Область научных интересов – разработка программных средств для оптимизации решения прикладных задач материаловедения.

А.В. Баглов

Окончил Белорусский государственный университет информатики и радиоэлектроники. Проводит научные исследования в рамках компьютерного моделирования электронной структуры перспективных материалов.

Л.С. Хорошко

Окончила Белорусский государственный университет информатики и радиоэлектроники. Проводит научные исследования в рамках разработки наноразмерных катализаторов фотостимулированных реакций и управляемого синтеза наноматериалов.

Аннотация. Успешное развитие науки в области синтеза и исследования современных и перспективных наноматериалов на сегодняшний день практически невозможно без использования вычислительных средств и компьютерной автоматизации. Более того, развитие нейронных сетей и программных интерфейсов (*machine learning*) для них позволяет не только исследовать, но и фактически предсказывать ключевые свойства исследуемых материалов – состав, стехиометрию, наличие и влияние дефектов кристаллов и др. Современные технологии и концепции *BigData* и *Advanced Analytics* в полной мере могут сыграть решающую роль в оптимизации подобных задач, сокращении времени обработки больших массивов данных, обеспечении возможности создания и использования глобальных информационных ресурсов для верификации. В данной статье рассматривается возможность использования компьютерного моделирования для автоматизации и повышения точности обработки результатов рентгеновской спектроскопии для изучения свойств наноматериалов, а также создания обобщенного электронного каталога/библиотеки различных образцов и применения методов машинного обучения.

Ключевые слова: рентгеновский дифракционный анализ, автоматизация, машинное обучение, облачные данные.

Введение.

Продуктивное слияние фундаментальной и прикладной науки и передовых информационных технологий весьма востребовано при ускоренных темпах современного научно-технического прогресса, а также с точки зрения укрепления позиций белорусской науки и образования на международной арене и развития концепции

импортозамещения [1].

Помимо предсказания свойств материалов [2], широкий ряд материаловедческих задач может быть решен с применением средств автоматизации, нейронных сетей и методов машинного обучения, что способствует снижению временных и трудовых затрат, что особенно актуально в рамках стратегии «Наука и технологии: 2018–2040» и внедрения концепции Университета 3.0 [3]. Также достаточно актуальным является создание глобальных баз данных, содержащих большое количество обработанных и верифицированных экспериментальных данных, а также удобство обращения с ними. Ряд существующих международных баз данных по структуре материалов являются как открытыми [4], так и предоставляемыми по лицензии [5], однако, конечная обработка результата поиска и сопоставление с собственным результатом осуществляется исследователем в «ручном» режиме.

Рассмотрим частный случай решения исследовательской задачи на примере обработки результатов одного из распространенных методов исследований свойств и структуры материалов – метода дифракции рентгеновских лучей (*XRD*, от англ. «*X-Ray Diffraction*»).

Постановка задачи.

Эффективность оценки характеристик материалов с применением обработки результатов *XRD* на сегодняшний день в значительной мере обусловлена скоростью и точностью определения положения и характеристик дифракционных пиков на получаемых дифрактограммах.

В традиционном представлении это требует от исследователя поиска и оценки параметров пика преимущественно в «ручном» режиме – определение точности позиционирования, интенсивности, *FWHM* (англ. «*full width at half maximum*» – полная ширина на половине амплитуды) для определения размера областей когерентного рассеяния и др. [5]. Значительно увеличивают время на оценку и интерпретацию результатов такие факторы как уровень шума на графике, базовая линия шума, разрывы данных и т.д. (Рисунок 1).

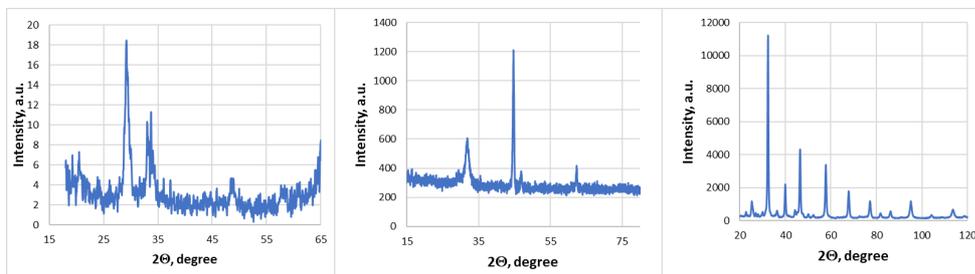


Рисунок 1. Примеры необработанных дифрактограмм различных материалов

Обработка и анализ сложных дифрактограмм может превышать по времени проведение самого эксперимента *XRD*. Также регулярно возникает проблема оперативного предоставления доступа к обработанным и верифицированным научным данным – результатам исследований, как коллегам по кафедре, так и всему научному сообществу в целом.

Данная проблема как правило «решается» в виде ручной пересылки писем или использования облачных хранилищ с предоставлением общего доступа ограниченному кругу лиц, что, в общем, оставляет вопрос многократной самостоятельной обработки результатов каждым исследователем в ручном режиме.

Для решения описанных проблем необходим комплексный подход, сочетающий в себе возможности оптимизации и ускорения процесса анализа дифрактограмм, сохранения результатов обработки и предоставления доступа к ним широкому кругу пользователей как в свободном режиме, так и с разграничением прав (к примеру, для постановки лабораторных и практических работ). Разработка проекта, описанного в данной статье, началась в 2023 г. на

кафедре физики твердого тела и нанотехнологий физического факультета БГУ.

Авторы проекта ставят перед собой цель предложить облачное решение в виде удобного пользовательского интерфейса (UI, от англ. «User Interface») для загрузки, оценки и анализа дифрактограмм (а в перспективе и иных спектроскопических данных – спектроскопии комбинационного рассеяния света, фотолюминесценции и др.), соединяющее пользователей по всему миру.

Методология и архитектура.

Для обеспечения актуального свойства максимальной совместимости со всем многообразием операционных систем и версий разрабатывается веб-приложение (SPA), которое взаимодействует с платформой посредством REST API, общий вид конфигурации платформы приведен на Рисунке 2.

Выбран облачный подход в виде провайдера AWS, который способен автоматически масштабировать нагрузку и является крайне высоко доступным сервисом. Сама платформа реализуется в виде serverless функций (AWS Lambda), для хранилища данных используется база данных Elasticsearch (AWS OpenSearch), которая поддерживает современные потребности BigData и способна масштабироваться горизонтально без особых усилий. Также рассматривается возможность интегрирования AWS RedShift в будущем.

Для работы приложению требуется выход в интернет, пользователи получают возможность заводить учетные записи, загружать научные данные, строить и публиковать необходимые графики. Интерфейс позволяет проводить автоматизированную обработку пользовательских или находящихся в базе дифрактограмм: удалить шум на графике, выровнять базовую линию шума, провести автоматический поиск пиков дифрактограммы и их позиций, а также провести аппроксимацию всех найденных или заданных пиков с применением функции Гаусса или Лоренца, для чего в программе реализуются методы математического моделирования. Также можно произвести автоматический расчет параметров пиков – амплитуда без учета шума при скорректированной базовой линии, FWHM – результаты отображаются на самом графике в виде наложений. Существует возможность добавления к графическому отображению дифрактограммы различных пользовательских пометок, комментариев, которые также могут быть сохранены для просмотра другими пользователями. Обработанные дифрактограммы сохраняются в облачной базе данных. Пользователи имеют возможность делиться графиками в режиме реального времени.

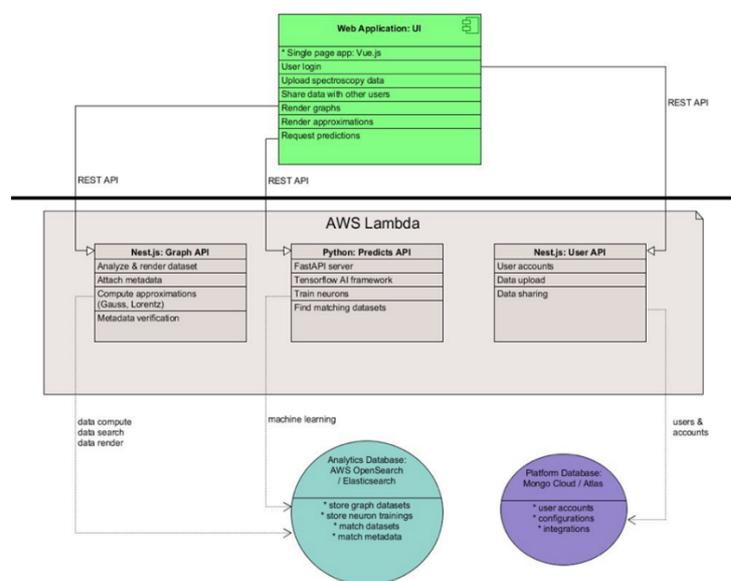


Рисунок 2. Общий вид конфигурации платформы (platform layout) разрабатываемого проекта

Ключевой особенностью приложения является возможность сопровождения каждого идентифицированного пика такими метаданными, как химический состав, фаза вещества, концентрация, и т.д. Это позволяет использовать для повышения точности и сокращения времени интерпретации такой современный подход, как машинное обучение. Так например обработанные дифрактограммы, имеющие полное и подтвержденное описание идентифицированных пиков и отмеченные как эталонные, используются в специальном модуле проекта (*Python*, фреймворк *Tensorflow AI*) для обучения нейронной сети, которая впоследствии способна предсказывать состав исследуемого материала с большой достоверностью на основании накопленных данных и проведенных циклов обучения нейронной сети. Достоверность прогнозов тем выше, чем больше выборка эталонов графиков в базе данных.

Результаты.

Методы математического моделирования, реализованные в приложении, показывают достаточно высокую точность в оценке распределения и позиционирования пиков в *XRD* спектрах. Эта точность может быть повышена за счет увеличения итераций (циклов) при построении аппроксимации пика. Такое увеличение хотя и влечет за собой дополнительное потребление вычислительной мощности, но способно значительно ускорять стадию автоматического анализа дифрактограмм. Применяемый подход машинного обучения позволяет практически мгновенно получать предположения о составе исследуемого наноматериала с помощью библиотеки накопленных пользовательских данных и обучаемой нейронной сети. Достоверность таких прогнозов искусственного интеллекта растет по мере накопления материала в облачной базе данных, а в качестве стартового набора могут быть использованы эталонные дифрактограммы, сгенерированные специализированными научными программами (например, *Vesta* и др.).

Следует отметить, что автоматическая оценка дифрактограмм, с точки зрения временных затрат на анализ полученных данных, значительно превосходит традиционную (ручную), а преимуществом разрабатываемого проекта по сравнению с существующими будет являться постоянное пополнение пользовательской базы данных и обучаемость. Выборочная оценка затраченного времени для исследования типовой дифрактограммы на примере постановки практической работы в учебной группе показывает, что в среднем для полного анализа дифрактограммы в ручном режиме требуется от 2 часов, с использованием демо-версии лицензируемого платного пакета *Match!* от 20 мин, то платформе требуется в среднем 12–18 секунд для проведения автоматического анализа такой же дифрактограммы, при этом в автоматическом режиме будет произведено устранение шумов, корректировка базовой линии и прочих артефактов.

Заключение.

Современные подходы в области *BigData* и *Advanced Analytics* показывают перспективные результаты в области исследования и предсказания свойств наноматериалов. Достоверность автоматизированного анализа с использованием нейросетей все еще является дискуссионным предметом в научном сообществе, и непрерывно растет, по мере совершенствования технологий в области *BigData* и *Machine Learning*, что открывает дальнейшие перспективы для дальнейшего углубления интеграции науки и информационных технологий.

Необходимо также отметить, что используемые вычислительные методологии не налагают особенных ограничений на источник и методики анализа материалов, которые могут быть автоматически обработаны. Так, например, существует возможность схожим образом автоматизировать обработку результатов, полученных с помощью таких методов, как Резерфордское обратное рассеяние, Рамановское рассеяние (комбинационное рассеяние света), электронная Оже-спектроскопия, люминесцентная спектроскопия и др., что подтверждает актуальность разрабатываемого продукта и перспективность проводимых исследований.

Список литературы

- [1] Перечень импортозамещающей продукции – Министерство экономики Республики Беларусь [Электронный ресурс]. URL: <https://economy.gov.by/ru/importoz-ru/> (Дата обращения: 28.03.2023).
- [2] Xie, T. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties / T. Xie, J. Grossman // *Physical Review Letters*. – 2018. – Vol. 120. – P. 145301-1–145301-6.
- [3] Стратегия «Наука и технологии: 2018–2040» [Электронный ресурс]. URL https://nasb.gov.by/congress2/strategy_2018-2040.pdf (Дата обращения: 20.03.2023).
- [4] Crystallography Open Database [Электронный ресурс]. URL: <http://www.crystallography.net/cod/> (Дата обращения: 20.03.2023).
- [5] Cambridge Structural Database System [Электронный ресурс]. URL: <https://software.chem.ucla.edu/CSD/> (Дата обращения: 20.03.2023).
- [6] Трушин, В.Н. Рентгеноский фазовый анализ поликристаллических материалов / В.Н. Трушин, П.В. Андреев, М.А. Фаддеев // Электронное учебно-методическое пособие. – Нижний Новгород: Нижегородский госуниверситет, 2012. – 89 с.

AUTOMATION OF PROCESSING THE RESULTS OF NANOMATERIALS STRUCTURE AND PROPERTIES STUDYING

N.A. Shymanski

Student of the Department of Solid State Physics and Nanotechnologies of the Faculty of Physics of BSU, Lead Software engineer of IDA Technologies

A.V. Baglov

Researcher in the R&D Lab of Energy-effective materials and technologies, senior lecturer of the Department of Solid State Physics of the Faculty of Physics of BSU

L.S. Khoroshko

Lead researcher in the R&D Lab of Energy-effective materials and technologies, associate professor of the Department of Solid State Physics of the Faculty of Physics of BSU, PhD (physics and mathematics)

*Department of Solid State Physics and Nanotechnologies
Faculty of Physics
Belarusian State University
E-mail: khoroshko@bsu.by*

Abstract. The successful development of science in the synthesis and study of modern and promising nanomaterials is practically impossible today without the integration of computing tools and computer automation. Moreover, the development of neural networks and software interfaces (machine learning) for them allows us to study as well as predict the main properties of studying materials, such as composition, stoichiometry, crystal defects presence, and effect, etc. Modern technologies and concepts of Big Data and Advanced Analytics can fully play a decisive role in optimizing such tasks, reducing the processing time of large data arrays, and enabling the creation and use of global information resources for verification. This article discusses the possibility of using computer simulation to automate and improve the accuracy of processing X-ray spectroscopy results for studying the properties of nanomaterials, creating a generalized electronic catalog/library of various samples, and applying machine learning methods.

Keywords: X-ray diffraction analysis, automation, machine learning, cloud data.