

План программы работ по освоению методов машинного обучения в ОФВЭ

К обсуждению.

Дата создания: 2023-12-12

Коррекция: 2023-12-15

Автор: А.Е. Шевель

Введение

Поскольку основной особенностью экспериментов ФВЭ является поиск редких (аномальных) многомерных объектов (событий) в огромном объёме многомерных данных, который часто превышает объёмы и сложность организации данных в других научных исследованиях. Эта особенность определяет необходимость разработки методов поиска редких событий. Одним из таких методов является технология машинного обучения, когда некоторая программа специальной архитектуры проходит обучение (тренировку), чтобы выявлять редкие события с уменьшением времени на вычисления. Хотя небезуспешные попытки использования методов машинного обучения в экспериментах по физике высоких энергий (ФВЭ) имели место уже в 80-х годах прошлого столетия, но последние 10 лет имел место весьма значительный прогресс в технологии машинного обучения. Важным результатом является тот факт, что эта технология стала применяться массовым порядком в различных экспериментах ФВЭ. Основной причиной массового применения упомянутых методов является ускорение получения результатов по сравнению с традиционными методами.

Имеется масса современных примеров использования методов машинного обучения в экспериментах ФВЭ, например, в детекторах [1]. Другие примеры использования машинного обучения приведены в приложении А.

По видимости следует планировать вычислительную установку машинного обучения, на которой сотрудники ОФВЭ могли бы получать опыт использования этой технологии: в каких отношениях она будет эффективна и каковы реальные ограничения. Следует рассматривать несколько стадий развития такой установки в ОФВЭ.

- Развёртывание минимального (базового) варианта, на котором можно получить первичный опыт для ряда сотрудников ОФВЭ.
- Развёртывание более сложной установки, которая может быть полезной для разработки или сопровождения крупного детектора, который будет использоваться в ОИЯИ или на другом крупном ускорителе.

Освоение методов машинного обучения касается всех научных сотрудников и инженеров ОФВЭ. *Каждому в какой-то степени следует понимать достоинства и ограничения систем машинного обучения.* Для этого необходимо организовать базовую для ОФВЭ компьютерную установку, которой смогут пользоваться некоторое число сотрудников. На начальном этапе такое число может быть оценено от 3 до 7 сотрудников ОФВЭ.

Минимальная установка для освоения машинного обучения

В минимальном варианте требуется сервер с соответствующей периферией под управлением Линукс. Вполне возможно, что сервер будет включён в микро-кластер ОФВЭ (rcfarm), т.е. доступ к нему будет достаточно обычным для сотрудников ОФВЭ.

Обычный сценарий использования технологии машинного обучения состоит в том, что используется некоторая готовая базовая модель, которая тренирована на довольно мощной установке типа супер-кластера. Чтобы такая модель была применима в конкретных задачах, её следует до-обучить (или дополнительно тренировать) на конкретных данных, соответствующих выбранной задаче. До-обучение или пере-обучение с целью оптимизации может потребоваться много раз. Именно для такого предполагается сервер ОФВЭ. После процесса оптимизации модель можно использовать на более скромных ресурсах (значительно более простых десктопах или ноутбуках).

Примерный минимальный состав сервера для до-обучения больших моделей

- CPU: ~ Intel i7 (3 GHz)
- GPU: на выбор (нужны от 1 до 3)
 - Видеокарта ASUS GeForce RTX 4090 24 ГБ (ROG STRIX GAMING 90YV0ID0-M0NA00) // Объем памяти 24 ГБ; Тип памяти GDDR6X; Частота графического процессора, МГц 2210; Тип шины PCI Express 4.0; Интерфейсы HDMI; Ширина шины памяти 384 бит; Поддерживаемые API DirectX, OpenGL; охлаждение Вентиляторное. // https://www.dns-shop.ru/product/3c1a5c449642ed20/videokarta-asus-geforce-rtx-4090-tuf-gaming-oc-edition-tuf-rtx4090-o24g-gaming/?utm_referrer=https%3A%2F%2Fwww.google.com%2F
Цена: 314092 Руб.
 - Графический процессор Nvidia Tesla A100 PCIe Gen4 (TCSA100M-80GB-PB) Tensor Cores 432, NVIDIA CUDA Cores

6912, 80 GB HBM2 with ECC, Memory bus width 5120-bit, Memory bandwidth 1,935 TB/s, Full Height, 300W, PCI Express 4.0 x16, Passive; // https://srv-trade.ru/catalog/videokarty_professionalnye_nvidia_i_amd/videokarty_ot_nvidia/graficheskiy_protssessor_serii_tesla/graficheskiy_protssessor_nvidia_tesla_a100_pcie_gen4_tcsa100m_80gb_pb_tensor_cores_432_nvidia_cuda_c.html // **Цена: 1434432 р.**

- Основная память: 128 GB (3600 MHz)
- Дисковая память: минимум 3 TB в виде трёх SSD по одному TB.
- Шина PCIe как минимум vers 4.
- Программное обеспечение:
 - Операционная система: Линукс
 - Модели машинного обучения (состав будет уточняться на основе обсуждаемых задач)

Масштаб затрат на оборудование:

- На GPU (от 315 т.р. до 4303 т.р. – разница в производительности несколько больше, чем на порядок)
- Системный блок:
 - Сервер Dell PowerEdge в комплектации: R750 (up to 16 x 2.5" NVMe U.2 HW Raid) rack 2U / 2 x Intel Xeon Gold 5315Y (3.2 Ghz, 8 cores, Cache 12MB, 140W, 2933 Mhz) / 2 x 64Gb PC4-25600(3200MHz) DDR4 ECC Registered DIMM / 3 x Dell 960GB, Data Center, NVMe, Read Intensive, U.2 / 2 x PERC H755N 8Gb RAID(0, 1, 5, 6, 10, 50, 60) 12Gb/s / no DVD / iDRAC 9 Enterprise / Embedded NIC 2x1Gb, No OCP Card / 1 x Power Supply, 800W, Hot-plug / no OS
 - Цена системного блока: 1541569 руб .

Оптимальный состав (системный блок + 2xGPU)

- Системный блок **1547865 руб**
- 2xGPU 2***1434432** **2868864 руб**
- Видео карта RTX 4090 **314092 руб**

Итого на оборудование:

1547865 + 2868864 + 314092 = 4730821 рублей

Примерный календарный план

Предполагается, что

- календарный план реализации оценивается в 24 месяца;
- общие трудозатраты оцениваются в 72 человеко-месяца.

Таким образом фонд зарплаты можно оценить как 7200 тыс. руб

Примерно 500 тыс. рублей могут быть оценены командировочные расходы за два года.

Итого: реалистическая оценка общей стоимости проекта

$4731 + 7200 + 500 = 12431$ тыс. рублей на два года

Ссылки

1. Stephen Roche, Quincy Bayer, Benjamin Carlson, William Ouligian, Pavel Serhiayenka, Joerg Stelzer, Tae Min Hong // Nanosecond anomaly detection with decision trees for high energy physics and real-time application to exotic Higgs decays // (submitted 7 Apr 2023) // <https://arxiv.org/abs/2304.03836>

Приложение А

Вопрос на perplexity.ai

Можно ли описать реальный эффект от использования искусственного интеллекта в конкретном эксперименте по физике высоких энергий?

Ответ Perplexity.ai

Да, можно описать реальный эффект от использования искусственного интеллекта (ИИ) в конкретном эксперименте по физике высоких энергий (ФВЭ). Например, эксперимент «Электрон-ионный коллайдер» (EIC), также известный как ePIC, представляет собой крупномасштабный эксперимент, который оптимизируется с помощью искусственного интеллекта. Детектор ePIC, который является частью сотрудничества более 170 учреждений по всему миру, будет идентифицировать частицы, образующиеся в результате столкновений, путем измерения их свойств, таких как скорость, масса и заряд. Система с использованием искусственного интеллекта, разрабатываемая для этого проекта, поможет оптимизировать конструкцию детектора. Ожидается, что этот проект сделает ePIC первым крупномасштабным экспериментом такого рода, который будет оптимизирован с помощью искусственного интеллекта [6].

Методы искусственного интеллекта и машинного обучения (ML) сыграли *важную роль в открытии бозона Хиггса*, идентификации происхождения частиц струй и используются в приложениях фильтрации событий в реальном времени [12]. Эти методы также используются при разработке более интеллектуальных детекторных систем, стратегий обработки и анализа данных, обеспечивая эффективное извлечение наиболее фундаментальной физической информации из множества данных, собранных на Большом адронном коллайдере (БАК) [4].

В будущем алгоритмы искусственного интеллекта, несомненно, займут ведущую роль в обработке и анализе данных в физике элементарных частиц. Например, Джессика Эскивель, физик-экспериментатор элементарных частиц, разрабатывает алгоритмы искусственного интеллекта для лучшего хранения и оптимизации луча в эксперименте Мюон g-2, который недавно показал несоответствие между предсказанными и измеренными значениями магнитного момента мюона [2].

Таким образом, ИИ оказывает значительное влияние на эксперименты НЕР: от оптимизации конструкции детектора до улучшения анализа данных и ускорения открытий.

Цитаты:

- [1] <https://worldscientific.com/worldscibooks/10.1142/12200>
- [2] <https://physicsworld.com/a/ai-and-particle-physics-a-powerful-partnership/>
- [3] https://www.academia.edu/84246499/Artificial_Intelligence_for_High_Energy_Physics_2022_english_
- [4] <https://youtube.com/watch?v=JVQbVGHHzs&t=1>
- [5] <https://www.frontiersin.org/articles/10.3389/frai.2022.999173>
- [6] <https://news.wm.edu/2023/09/13/accelerating-experimental-nuclear-physics-with-ai/>
- [7] <https://iopscience.iop.org/article/10.1088/1742-6596/1085/2/022008/pdf>
- [8] <https://ui.adsabs.harvard.edu/abs/2021APS..APRH02001T/abstract>
- [9] <https://moderndiplomacy.eu/2023/09/28/artificial-intelligence-and-advances-in-physics-in-the-field-of-gravitational-waves-i/>
- [10] <https://physics.aps.org/articles/v16/166>
- [11] <https://www.turing.ac.uk/events/nature-reviews-physics-machine-learning-theoretical-and-experimental-high-energy-physics>
- [12] <https://arxiv.org/pdf/2212.05081.pdf>
- [13] <https://www.sciencedirect.com/science/article/abs/pii/S0168900203006077>
- [14] <https://researchfunding.duke.edu/artificial-intelligence-research-high-energy-physics>
- [15] https://www.physics.purdue.edu/news/2022/1201_jung_liu_doe.html
- [16] <https://www.diva-portal.org/smash/get/diva2:1597833/FULLTEXT02>
- [17] <https://arxiv.org/pdf/2206.09645.pdf>

И ещё

Искусственный интеллект (ИИ) использовался различными способами для анализа данных физики высоких энергий. Вот некоторые примеры:

1. Большой адронный коллайдер (БАК) в ЦЕРН: искусственный интеллект использовался для улучшения и автоматизации обработки данных, количественной оценки систематических неопределенностей в моделировании на основе данных и повышения чувствительности к непредвиденным сигналам новой физики. Алгоритмы машинного обучения были разработаны для создания высокоточных классификаторов, построенных непосредственно на данных низкого уровня, автоматизирующих стандартную реконструкцию и избегающих слепых зон, присущих традиционному подходу [2]. ИИ также использовался для разработки управляемой данными триггерной системы для БАК, которая помогает фильтровать и анализировать огромные объемы собираемых данных [6].

2. Идентификация частиц: ИИ использовался для идентификации неуловимых сигнатур частиц среди зашумленных данных, повышая точность идентификации частиц в таких экспериментах, как БАК [5]. Метод квантового машинного обучения был применен для классификации так называемых b -джетов, струй, возникающих из b -кварков в результате протон-протонных столкновений в эксперименте LHCb [7].

3. Обнаружение аномалий. Исследовательская группа БАК под руководством Изобель Охалво из Принстона разрабатывает «алгоритм обнаружения аномалий», который использует машинное обучение для выявления событий столкновений, которые являются особенно редкими или необычными. Этот алгоритм станет частью серии триггерных систем, подключенных к одному из основных детекторов БАК — детектору компактного мюонного соленоида (CMS) [12].

4. Анализ данных и моделирование событий: искусственный интеллект применялся в нескольких аспектах физики высоких энергий, особенно в экспериментах на БАКе, в триггерных системах и системах сбора данных, вычислительных задачах, моделировании событий, реконструкции, работе детекторов и физическом анализе. [9].

Эти примеры иллюстрируют, как методы искусственного интеллекта и машинного обучения используются для анализа и интерпретации данных физики высоких энергий, что приводит к более эффективным и точным результатам.

Цитаты:

- [1] <https://worldscientific.com/worldscibooks/10.1142/12200>
- [2] <https://www.cmu.edu/ai-physics-institute/research/particle.html>
- [3] <https://sparks.cern/index.php/ai-cern>
- [4] <https://www.frontiersin.org/journals/big-data/sections/big-data-and-ai-in-high-energy-physics>
- [5] <https://rescale.com/blog/ai-physics/>
- [6] <https://datascience.uchicago.edu/research/a-data-driven-trigger-system-for-the-large-hadron-collider/>
- [7] <https://www.nature.com/articles/s41534-021-00443-w>
- [8] <https://www.usparticlephysics.org/brochure/particle-physicists-advance-artificial-intelligence/>
- [9] <https://www.innovationnewsnetwork.com/big-data-challenge-large-hadron-collider/11359/>
- [10] <https://www.oecd-ilibrary.org/sites/724b14a6-en/index.html?itemId=%2Fcontent%2Fcomponent%2F724b14a6-en>
- [11] https://www.symmetrymagazine.org/article/machine-learning-proliferates-in-particle-physics?language_content_entity=und
- [12] <https://research.princeton.edu/news/what-universe-made-ai-could-help-us-find-out>
- [13] <https://physicsworld.com/a/ai-and-particle-physics-a-powerful-partnership/>
- [14] <https://www.nature.com/articles/d41586-018-05084-2>
- [15] <https://youtube.com/watch?v=JVQbVGHHrzs>
- [16] <https://www.scientificamerican.com/article/artificial-intelligence-called-in-to-tackle-lhc-data-deluge/>

- [17]<https://www.frontiersin.org/journals/artificial-intelligence/sections/big-data-and-ai-in-high-energy-physics/about>
- [18]https://nucleus.iaea.org/sites/ai4atoms/WGAI4PHY%20Documents/4-2_Valentino.pdf
- [19] <https://iopscience.iop.org/article/10.1088/1742-6596/1085/2/022008/pdf>
- [20]<https://home.cern/news/news/physics/speeding-machine-learning-particle-physics>