



**МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
имени М.В. ЛОМОНОСОВА**



**НАУЧНО-ИССЛЕДОВАТЕЛЬСКИЙ ИНСТИТУТ
ЯДЕРНОЙ ФИЗИКИ
имени Д.В. СКОБЕЛЬЦИНА**

РАЗВИТИЕ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ И ТЕЛЕКОММУНИКАЦИЙ

Отчет за 2022 год

*Москва
2023*

УДК 539.165:537,591
БИК 22.38

РАЗВИТИЕ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ И ТЕЛЕКОММУНИКАЦИЙ

Отчет за 2022 год

Под редакцией член-корр. РАН Э.Э. Бооса, Е.А. Сигаевой

РАЗВИТИЕ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ И ТЕЛЕКОММУНИКАЦИЙ.
Отчет за 2022 год /Под редакцией член-корр.РАН Э.Э. Бооса,
Е.А. Сигаевой; Московский государственный университет имени
М.В. Ломоносова, Научно-исследовательский институт ядерной физики
имени Д.В. Скобельцына. — М., 2023.

В сборнике содержится краткое описание полученных в НИИЯФ МГУ в 2022 г. результатов в области информационных технологий и телекоммуникаций, составленное отделом научной информации на основе данных, предоставляемых информационно-аналитической системой «Наука — МГУ» и опубликованных в отечественной и зарубежной литературе.

**К 270-ЛЕТИЮ СО ДНЯ ОСНОВАНИЯ
МОСКОВСКОГО ГОСУДАРСТВЕННОГО УНИВЕРСИТЕТА
ИМЕНИ М.В. ЛОМОНОСОВА**



СОДЕРЖАНИЕ

РАЗВИТИЕ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ И ТЕЛЕКОММУНИКАЦИЙ. КРАТКАЯ СПРАВКА	5
ОСНОВНЫЕ РЕЗУЛЬТАТЫ, ПОЛУЧЕННЫЕ В НИИЯФ МГУ В 2022 г. В ОБЛАСТИ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ И ТЕЛЕКОММУНИКАЦИЙ	6
Разработка нейросетевых методов обработки данных физического эксперимента	6
Разработка и исследование распределенных информационно-вычислитель- ных систем для обработки и анализа больших данных в физических экспе- риментах	9
СОБЫТИЯ 2022 года	11
6-ая Международная школа-семинар по глубокому обучению в вычислитель- ной физике	11

РАЗВИТИЕ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ И ТЕЛЕКОММУНИКАЦИЙ. КРАТКАЯ СПРАВКА

В НИИЯФ МГУ ведутся работы по созданию и использованию новых информационных технологий.

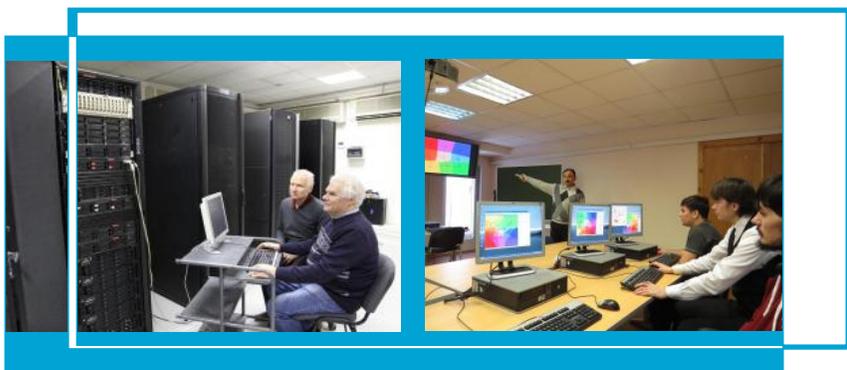
Институт является одним из первопроходцев в России в области построения инфраструктуры распределенных вычислений ГРИД, которые внесли огромный вклад в открытие бозона Хиггса и других результатов последних лет. Основываясь на опыте, полученном в процессе сотрудничества с учеными ЦЕРН, на вычислительных ресурсах ЛАВФВЭ НИИЯФ МГУ развернута инфраструктура, которая обеспечивает хранение, обработку и анализ данных экспериментов в области гамма-астрономии TAIGA/TUNKA, что позволило получить ряд новых интересных результатов. Эту же инфраструктуру планируется использовать для анализа данных экспериментов JUNO (КНР) и NICA (Россия).

Также в НИИЯФ МГУ развиваются нейросетевые и другие информационные технологии, связанные с машинным обучением, для решения научно-исследовательских физических задач. На основе полученного в этой области опыта созданы специализированные учебные курсы, на которых реализуется программа повышения квалификации по машинному обучению, искусственным нейронным сетям и генетическим алгоритмам для специалистов различных отраслей.

Тематический план научно-исследовательских работ НИИЯФ МГУ в 2022 г. включал в себя 2 темы исследований в области информационных технологий и телекоммуникаций:

- «Разработка нейросетевых методов обработки данных физического эксперимента»;
- «Разработка и исследование распределенных информационно-вычислительных систем для обработки и анализа больших данных в физических экспериментах»

а)



б)

Работа по настройке ГРИД-кластера в НИИЯФ МГУ (а)

Занятия на курсах повышения квалификации по искусственным нейронным сетям и генетическим алгоритмам (б)

ОСНОВНЫЕ РЕЗУЛЬТАТЫ, ПОЛУЧЕННЫЕ В НИИЯФ МГУ В 2022 г. В ОБЛАСТИ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ И ТЕЛЕКОММУНИКАЦИЙ

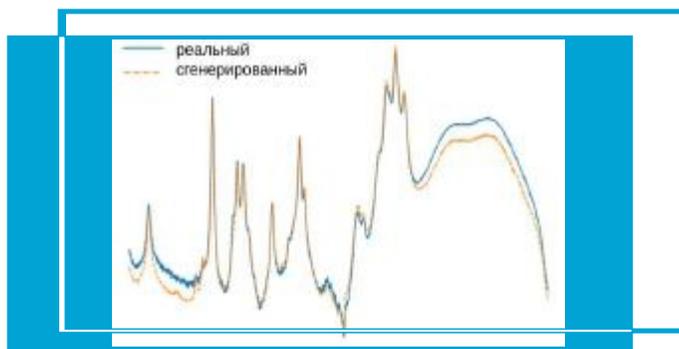
Разработка нейросетевых методов обработки данных физического эксперимента

Решение многопараметрических обратных задач

Продемонстрирована возможность повышения качества решения обратных задач (ОЗ) путем нейросетевой генерации и использования дополнительных модельных примеров. В качестве такой задачи выбрана многопараметрическая ОЗ по определению концентраций компонент в многокомпонентных растворах. Для генерации модельных спектров использовались обусловленные вариационные автоэнкодеры (ВАЭ) – специальный вид генеративных сетей, с заданием условий в латентном пространстве ВАЭ, с использованием кластеризации и моделей проекций на латентные структуры. Подача на генератор вектора-условия позволила генерировать примеры, соответствующие определенным концентрациям компонентов. А добавление генерированных примеров к тренировочному набору из реальных примеров – получить заметное (до 36%, обычно 15-17%) снижение погрешности решения.

Для решения ОЗ разведочной геофизики (подход «от модели») путем расчета прямой задачи были получены уникальные 2D- и 3D-массивы данных для слоистых сред с переменной толщиной слоёв, для трёх видов разведочной геофизики; гравиразведка, магниторазведка, электроразведка (магнитотеллурика (по 26 000 – 30 000 примеров). Решение ОЗ в режиме комплексирования данных (одновременного использования данных разных физических методов) показало, что комплексирование позволяет существенно снизить погрешность решения ОЗ.

Для решения ОЗ оптической спектроскопии (подход от «эксперимента») путем физического эксперимента были получены уникальные массивы данных многокомпонентных растворов солей тяжелых металлов для диагностики жидких технологических сред (для 3762 растворов на дистиллированной воде и 1000 растворов на воде из московских рек), для рН-метрии и трех видов оптической спектроскопии (оптического поглощения, ИК-поглощения, комбинационного рассеяния). Показано, что эффективность комплексирования данных различных видов спектроскопии зависит от того, как соотносятся между собой погрешности решения для данных отдельных методов.

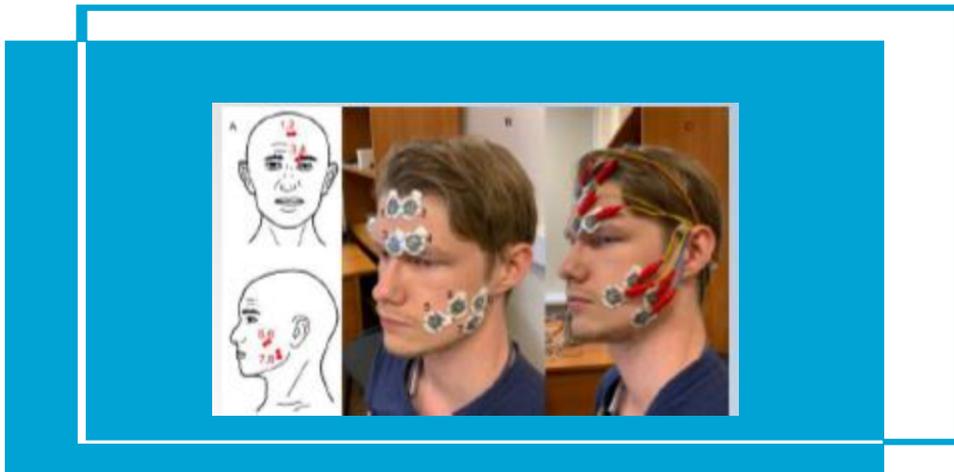


Сравнение сгенерированного спектра с реальным

все индивидуумы имеют одинаковые свойства. В гендерном ГА (ГГА) индивидуумы делятся на два пола (гендера). Внутри одного гендера идет отбор, между разными – скрещивание. Для мужских индивидуумов характерна более высокая вероятность мутации, а для женских – ограничение на число скрещиваний. В результате исследований получено, что ГГА сходится дольше (популяция вырождается позже). Помимо этого, ГГА справляется с многоэкстремальными функциями с явно выраженным экстремумом лучше, чем ГА.

Обработка данных в когнитивных исследованиях

Рассматривались данные электромиографии (ЭМГ) с трех мышц лица. Испытуемым демонстрировали стимулы эмоций. Рассматривались два варианта пространства эмоций: 1) валентность – возбуждение – доминантность; 2) радость – грусть – гнев – удивление – испуг – отвращение. Строились отображения из пространства сигналов ЭМГ мышц в пространстве эмоций и их первых главных компонент. Далее проводилось выделение из сигналов специального набора признаков (погружение, стат. показатели, скользящие средние) и отбрасывание наименее существенных признаков. Показано, что присущая размерность пространства эмоций – 3. Все компоненты этого пространства можно восстановить с достаточной точностью из сигналов мышц с помощью машинного обучения. Однако оказалось, что разные эмоции восстанавливаются с разной эффективностью – например, радость восстанавливается хорошо, а удивление и испуг – плохо.



Прогноз петротипа и фильтрационно-емкостных свойств породы

Необходимо научиться определять свойства горной породы по результатам пробного бурения, желательно – по шламу. Система автоматизированной минералогии на основании электронной микроскопии позволяет определять минеральный состав породы. На основании минерального состава породы методами машинного обучения можно определять: петротип породы, харак-

теризующий ее свойства, свойство коллектор/ неколлектор, фильтрационно-емкостные свойства – пористость и проницаемость. Было выполнено сравнение результатов кластеризации данных с разделением по петротипам и сравнение результатов для керна, имитации шлама и шлама. Разработаны рекомендации по совершенствованию системы. Реализован программный комплекс для решения перечисленных задач.

Преподавание в области машинного обучения

Созданы и реализуются 2 курса: «Машинное обучение в физике» (факультатив для студентов) (осень) и «Машинное обучение. Искусственные нейронные сети и генетические алгоритмы» (курсы повышения квалификации и факультатив для студентов). На курсах изучаются искусственные нейронные сети и другие алгоритмы машинного обучения; основы предобработки данных; генетические алгоритмы, генетическое программирование; нечеткая логика, метод группового учета аргументов. Курсы базируются на современном инструментарии для практической работы по анализу данных. Практические занятия ведутся с использованием языков R и Python. Слушатели выполняют самостоятельные работы под руководством преподавателя. Для студентов МГУ обучение бесплатное.

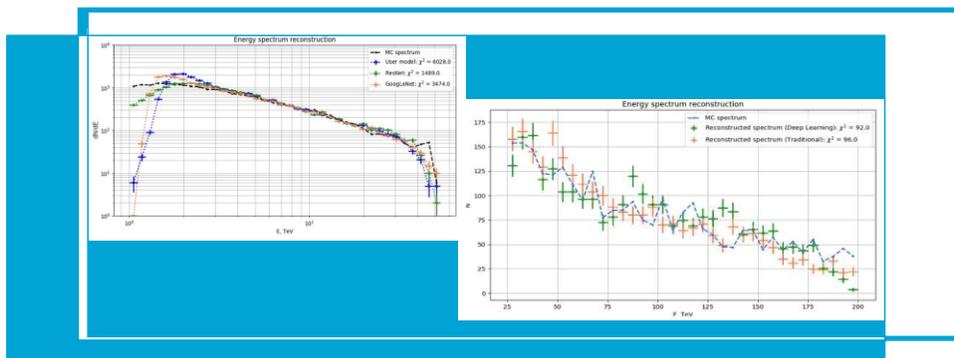


Разработка и исследование распределенных информационно-вычислительных систем для обработки и анализа больших данных в физических экспериментах

Применение методов машинного обучения в гамма-астрономии

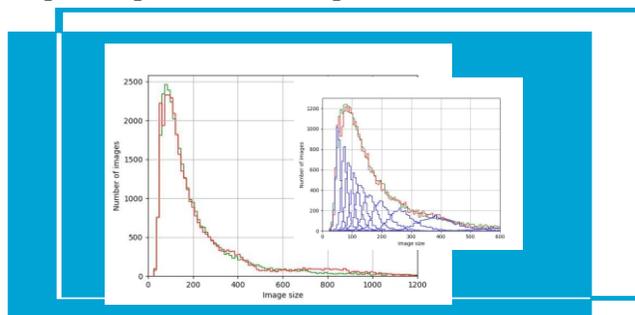
В НИИЯФ МГУ проводится подготовка к внедрению методов машинного обучения в реальную практику обработки данных астрофизического комплекса TAIGA. В настоящее время этот комплекс регистрирует черенковское излучение широких атмосферных ливней (ШАЛ) с помощью 3 телескопов, расположенных в долине Тунка (Бурятия). Для обработки получаемых данных традиционно используются так называемые параметры Хилласа, извлеченные из событий и характеристик ШАЛ. В 2022 году изучались возможности сверхточных нейронных сетей для классификации типа первичных частиц ШАЛ. В качестве таких сетей применялись сети User CNN, ResNet и GoogLeNet.

Для корректного сравнения возможностей этих сетей применялось примерно одинаковое количество параметров около – 2 млн. Было выполнено сравнение реконструированных спектров гамма-квантов. Сравнение проводилось на данных метода моделирования Монте-Карло (МК) в диапазоне энергии γ -квантов 25–200 ТэВ. В качестве искусственной нейронной сети использовалась сеть с двумя входами (стереорежим). Исследование показало, что использование сверточных нейронных сетей для классификации типа первичных частиц дает очень хорошие результаты. Получено хорошее согласие результатов применения сверточных нейронных сетей для восстановления спектров гамма-частиц с традиционными методами. Заметное улучшение дает одновременное использование данных с нескольких телескопов (стереорежим).



Энергетические спектры событий ШАЛ, реконструированные с помощью сверточных нейронных сетей

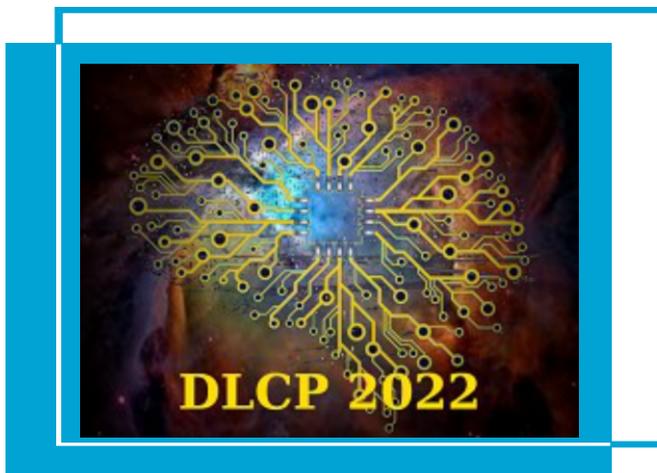
Генеративно-сопоставительные сети (GAN) предназначены для генерации изображений, отличающихся от реальных; они используются как быстрая замена МК-генераторов. В 2022 г. было выполнено моделирование γ -событий с помощью условной GAN. Для обучения сети весь набор событий был разделен на 10 наборов равных по количеству событий (примерно 3500 событий в каждом наборе). Каждый набор рассматривался как отдельный класс, номер которого дополнительно передавался в сеть. При генерации событий моделировалось одинаковое количество событий во всех классах. Скорость генерации составила примерно 5000 событий в сек. Генеративные нейронные сети продемонстрировали свой большой потенциал в качестве замены МК-генераторов, их использование ускорило процесс моделирования событий в сотни раз.



Результат работы генеративно-сопоставительной нейронной сети. Зеленая линия – МК-распределение, синяя – распределение отдельных классов (с 1 по 10), красная – суммарное распределение

СОБЫТИЯ 2022 г.

6-ая Международная школа-семинар по глубокому обучению в вычислительной физике



С 6 по 8 июля в Объединенном институте ядерных исследований (ОИЯИ, Дубна) состоялась 6-ая Международная школа-семинар по глубокому обучению в вычислительной физике (DLCP-2022). Мероприятие было организовано Лабораторией информационных технологий имени Мещерякова (МЛИТ) ОИЯИ и НИИЯФ МГУ. Оно прошло в смешанном формате: очном и дистанционном. Школа-семинар была посвящена использованию машинного обучения в астрофизике элементарных частиц, физике высоких энергий и других областях. Основными темами были различные приложения искусственных нейронных сетей к физическим задачам, а также разработка новых современных методов машинного обучения для анализа различных научных данных, в том числе больших данных.



Для заметок

