

Форма «Т». Титульный лист заявки в Российский научный фонд
Конкурс 2021 года «Проведение фундаментальных научных исследований и
поисковых научных исследований малыми отдельными научными группами»

Название проекта Моделирование выборок случайных событий с учетом априорной информации в астрофизических экспериментах методами машинного обучения	Номер проекта 22-21-00442	
	Отрасль знания: 01	
	Основной код классификатора: 01-202 Дополнительные коды классификатора: 01-218 01-726	
	Код ГРНТИ 50.41.25	
Фамилия, имя, отчество (при наличии) руководителя проекта: Крюков Александр Павлович	Контактные телефон и e-mail руководителя проекта: +74959393156, +79163630991, kryukov@theory.sinp.msu.ru	
Полное и сокращенное наименование организации, через которую должно осуществляться финансирование проекта: Федеральное государственное бюджетное образовательное учреждение высшего образования «Московский государственный университет имени М.В.Ломоносова» Московский государственный университет имени М.В.Ломоносова, Московский университет или МГУ		
Объем финансирования проекта в 2022 г.: 1500 тыс. руб.	Год начала проекта: 2022	Год окончания проекта: 2023
Гарантирую, что при подготовке заявки не были нарушены авторские и иные права третьих лиц и/или имеется согласие правообладателей на представление в Фонд материалов и их использование Фондом для проведения экспертизы и для обнародования (в виде аннотаций заявок).		
Подпись руководителя проекта _____ /А.П. Крюков/		Дата регистрации заявки 10.06.2021 г.
Подпись руководителя организации* * Либо уполномоченного представителя, действующего на основании доверенности или распорядительного документа. В случае подписания формы уполномоченным представителем организации (в т.ч. - руководителем филиала) к печатному экземпляру заявки <u>прилагается копия распорядительного документа или доверенности</u> , заверенная печатью организации. Непредставление копии распорядительного документа или доверенности в случае подписания формы уполномоченным представителем организации, а также отсутствие расшифровки подписи, является основанием недопуска заявки к конкурсу. _____ / _____ /		
Печать (при наличии) организации		

Форма 1. Сведения о проекте

1.1. Название проекта

на русском языке

Моделирование выборок случайных событий с учетом априорной информации в астрофизических экспериментах методами машинного обучения

на английском языке

Modeling samples of random events with the account of prior knowledge in astrophysical experiments using machine learning methods

1.2. Приоритетное направление развития науки, технологий и техники в Российской Федерации, критическая технология

Указывается согласно перечню (Указ Президента Российской Федерации от 7 июля 2011 года №899) в случае, если тематика проекта может быть отнесена к одному из приоритетных направлений, а также может внести вклад в развитие критических технологий Российской Федерации.

3. Информационно-телекоммуникационные системы.

18. Технологии и программное обеспечение распределенных и высокопроизводительных вычислительных систем.

Направление из Стратегии научно-технологического развития Российской Федерации (утверждена Указом Президента Российской Федерации от 1 декабря 2016 г. № 642 «О Стратегии научно-технологического развития Российской Федерации») (при наличии)

1.3. Ключевые слова (приводится не более 15 терминов)

на русском языке

машинное обучение, моделирование физических явлений, генеративные нейронные сети, Монте-Карло

на английском языке

machine learning, modeling of physics phenomena, generative neural networks, Monte Carlo

1.4. Аннотация проекта (объемом не более 2 стр.; в том числе кратко – актуальность решения указанной выше научной проблемы и научная новизна)

Данная информация может быть опубликована на сайте Фонда в информационно-телекоммуникационной сети «Интернет».

на русском языке

Научной проблемой, на решение которой направлен проект, является исследование и разработка методов машинного обучения для анализа моделей физических явлений с использованием выборок случайных состояний (семплов). До последнего времени практически единственным подходом, позволяющим численно моделировать очень сложные физические явления, являлся подход на основе методов Монте-Карло. Однако использование реалистичных физических моделей на основе метода Монте-Карло требуют весьма больших временных и ресурсных затрат. Одним из альтернативных и весьма перспективных методов является использование методов машинного обучения для генерации выборок случайных векторов.

Отличительной особенностью моделирования физических явлений, особенно в физике элементарных частиц и космических лучей, является необходимость воспроизвести состояния системы с учетом ограничений, определяемых природой физического явления. Априорные знания могут иметь форму логических правил, алгебраических или дифференциальных уравнений (например законы сохранения или уравнения связей между параметрами или степенями свободы, описывающими изучаемые явления), наличия симметрий относительно дискретных или непрерывных групп входных/выходных данных, а также заранее известные вероятностные соотношения между данными. В методе Монте-Карло такие ограничения вводятся на уровне математической модели. Для методов машинного обучения, основанных на принципе обучения, такие ограничения воспроизводятся с трудом. Особенно это касается различных статистических распределений по совокупности состояний. Эти проблемы можно решить путем дополнительного включения априорных знаний о таких ограничениях в процесс обучения. В этом состоит новизна предлагаемого подхода к решению общей проблемы – эффективного генерирования выборок случайных состояний (семплов) с учетом априорных знаний. Практическая проверка разрабатываемых подходов будет применена для анализа моделей физических явлений в астрофизике частиц.

Конкретной задачей данного проекта в рамках общей проблемы является исследование и разработка методов

глубокого обучения генерации выборок многомерных случайных векторов в пространствах с большим числом измерений, с учетом априорной детерминистской и вероятностной информации о генерируемых событиях и их выборках для моделирования сложных физических явлений на примере экспериментов в области наземной гамма-астрономии. Наземная гамма-астрономия изучает излучение гамма-квантов галактического и внегалактического происхождения с помощью специально разработанных установок, так называемых атмосферных черенковских гамма-телескопов (Imaging Atmospheric (or Air) Cherenkov Telescope, IACT). С помощью этого метода гамма-излучение регистрируется на земле оптически как черенковский свет, порождаемый обширными ливнями вторичных частиц, когда гамма-кванты очень высокой энергии попадают в атмосферу. Гамма-кванты таких энергий составляют лишь малую долю (меньше одной десяти тысячной) потока космических лучей, состоящего, в основном, из протонов. Для анализа данных, планирования эксперимента и других задач необходимо проводить моделирование, результатом которого должны стать сотни тысяч и миллионы событий, включая протонные события, являющиеся фоном. На разработку эффективных алгоритмов машинного обучения для генерации семплов и направлен предложенный проект. Решение данной проблемы является весьма важным для гамма-астрономии, поскольку благодаря отсутствию электрического заряда, гамма-кванты несут информацию об их сверхудаленных источниках, в которых происходят экзотические и экстремальные процессы во Вселенной.

Интеграция в процесс обучения априорных, заранее известных знаний позволит заметно повысить качество полученных выборок, а следовательно и качество моделирования событий, в частности, событий широких атмосферных ливней, регистрируемых черенковскими телескопами. Практическими задачами проекта будут: исследование существующих и разработка новых методов машинного обучения для генерации случайных векторов с учетом априорной информации; разработка алгоритмов на их основе и их программная реализация; исследование, сравнение, выбор наилучшего (или лучших) методов; сравнение с существующими методами. В ходе осуществления проекта будет проведен полный цикл исследований и решения поставленных задач — от теоретической разработки подхода до практической программной реализации.

Помимо базового примера из области астрофизических экспериментов, генераторы событий на основе методов машинного обучения найдут важные применения в других научных и прикладных областях, связанных с анализом сложных и многомерных вероятностных явлений. Например, в области гидродинамики при исследовании процессов, связанных с турбулентностью, в метеорологии, в исследованиях поведения больших групп людей или животных в критических обстоятельствах и других. Из-за взрывного роста объема разнообразных данных, которые требуют анализа, в том числе с помощью существующих и новых методов машинного обучения, проблема генерации выборок случайных состояний (семплов событий), соответствующих априорным знаниям, в случае пространств с большим числом измерений является весьма актуальной и важной.

на английском языке

The scientific problem to be solved under the project is the research and development of deep machine learning methods for analyzing models of physical phenomena using samples of random states. Until recently, practically the only approach that allows one to numerically simulate very complex physical phenomena was the approach based on Monte Carlo methods. However, the use of realistic physical models based on the Monte Carlo method is very time and resource intensive. One of the alternative and very promising methods is the use of machine learning methods to generate samples of random vectors.

A distinctive feature of modeling physical phenomena, especially in the physics of elementary particles and cosmic rays, is the need to reproduce the states of the system, taking into account the constraints determined by the nature of the physical phenomenon. Prior knowledge can take the form of logical rules, algebraic or differential equations (for example, conservation laws or equations of relationships between parameters or degrees of freedom that describe the phenomena under study), the presence of symmetries with respect to discrete or continuous groups of input/output data, as well as known beforehand probabilistic relationships among data. In the Monte Carlo methods, a prior knowledge is introduced at the level of the mathematical model. For machine learning methods, such constraints are difficult to reproduce. This is especially true for various statistical distributions over a set of states. These problems can be solved by additionally incorporating prior knowledge of such constraints into the learning process. This is the novelty of the proposed approach to solving the general problem, namely the efficient generation of samples of random states taking into account prior knowledge. Practical verification of the developed approaches will be carried out by applying them to analysis of models of physical phenomena in astroparticle physics.

The specific task of this project within the framework of the general problem is to research and develop methods of deep learning for generating samples of multidimensional random vectors in spaces with a large number of dimensions, taking into account a priori deterministic and probabilistic information about generated events and their samples for modeling complex physical phenomena using the example of experiments in the field of terrestrial gamma-ray astronomy. Terrestrial gamma-ray astronomy studies the emission of gamma quanta of galactic and extragalactic origin using specially designed installations, the so-called atmospheric (or air) Cherenkov gamma telescopes (IACT). With this method, gamma rays are optically detected on the ground as Cherenkov light, generated by extensive showers of secondary particles when very high energy gamma rays enter the atmosphere. Gamma quanta of such energies make up only a small fraction (less than one ten-thousandth) of the cosmic ray flux, which consists mainly of protons. For data analysis, experiment planning and other tasks, it is necessary to carry out simulations, the result of which should be hundreds of thousands and even millions of events, including proton events that are the background. The proposed project is aimed at developing effective machine learning algorithms for generating samples. The solution to this problem is very important for gamma astronomy, since due to the absence of an electric charge, gamma quanta carry information about their ultra-distant sources, in which exotic and extreme processes in the Universe take place.

Integration of a priori, previously known knowledge into the learning process will significantly improve the quality of the obtained samples, and, consequently, the quality of event modeling, in particular, events of extensive air showers recorded by Cherenkov telescopes. The practical tasks of the project will be: research of existing and development of new methods of deep machine learning for generating random vectors taking into account prior informations; development of algorithms based on them and their software implementation; research, comparison, selection of the best methods; comparison with existing methods. During the implementation of the project, a full cycle of research and solutions of the assigned tasks will be carried out, from theoretical development of the approach to practical software implementation.

In addition to the basic example from the field of astrophysical experiments, event generators based on deep machine learning methods will find important applications in other scientific and applied areas related to the analysis of complex and multidimensional probabilistic phenomena. For example, in the field of hydrodynamics for the study of processes associated with turbulence, in meteorology, in the study of the behavior of large groups of people or animals in critical circumstances, and others. Due to the explosive growth in the amount of various data that require analysis, including that based on existing and new methods of deep machine learning, the problem of generating samples of random states corresponding to prior knowledge, in the case of spaces with a large number of dimensions, is very relevant and important.

1.5. Ожидаемые результаты и их значимость (указываются результаты, их значимость для развития новой научной тематики)

Данная информация может быть опубликована на сайте Фонда в информационно-телекоммуникационной сети «Интернет».

на русском языке

Основными результатами данного проекта будут методы интеллектуального моделирования данных в астрофизических экспериментах на основе машинного обучения. К ним относятся методы быстрого моделирования событий регистрации широких атмосферных ливней черенковскими телескопами с использованием генеративных нейронных сетей. Их научная значимость состоит в том, что использование таких методов позволит заменить сложные и ресурсоемкие методы Монте-Карло, существенно ускорить процесс моделирования, что, в свою очередь, позволит существенно продвинуться в понимании фундаментальных процессов, происходящих во Вселенной. Разрабатываемые методы могут найти применение не только в научных исследованиях, но и в инженерных науках, в экономике, везде, где используются методы Монте-Карло для моделирования различных явлений и необходимо учитывать априорную информацию, присущую этим явлениям. Конкретно, в рамках проекта будут получены следующие основные результаты:

- методы и алгоритмы на основе машинного обучения генерации случайных выборок многомерных векторов с учетом априорных ограничений, в том числе, статистических характеристик полученных наборов (семплов);
- программная адаптация существующих и реализация новых методов и алгоритмов генерации случайных выборок;
- результаты вычислительных экспериментов по сравнению Монте-Карло генераторов и генераторов основанных на машинном обучении с учетом априорной информации на примере моделирования событий в гамма-астрономии.

на английском языке

The main results of this project will be methods of intelligent data modeling in astrophysical experiments based on deep machine learning. These include methods for fast modeling of events of registration of large air showers by Cherenkov telescopes using generative neural networks. Their scientific significance lies in the fact that the use of such methods will allow replacing complex and resource-intensive Monte Carlo methods, significantly speeding up the modeling process,

which, in turn, will make it possible to significantly advance in understanding the fundamental processes occurring in the Universe. The developed methods can find application not only in scientific research, but also in engineering sciences, in economics, wherever Monte Carlo methods are used to model various phenomena and it is necessary to take into account a prior information inherent in these phenomena.

Specifically, within the framework of the project, the following main results will be obtained:

- methods and algorithms based on deep machine learning for generating random samples of multidimensional vectors, taking into account prior constraints, including the statistical characteristics of the obtained sets (samples);
- software adaptation of existing and implementation of new methods and algorithms for generating random samples;
- the results of computational experiments comparing Monte Carlo generators and generators based on machine learning, taking into account a prior information on the example of event modeling in gamma astronomy.

1.6. В состав научного коллектива (в т.ч. с учетом руководителя проекта) будут входить (указывается планируемое количество исполнителей в течение всего срока реализации проекта):

Несоответствие состава научного коллектива (в том числе отсутствие информации в соответствующих полях формы) требованиям пункта 12 конкурсной документации является основанием недопуска заявки к конкурсу.

4 исполнителей проекта (включая руководителя),

В соответствии с требованиями пункта 12 конкурсной документации от 2 до 4 человек, вне зависимости от того, в трудовых или гражданско-правовых отношениях исполнители состоят с организацией.

В том числе:

- 2** исполнителей в возрасте до 39 лет включительно;
- 0** аспирантов (адъюнктов, интернов, ординаторов) очной формы обучения;
- 2** студентов очной формы обучения.

1.7. Планируемый состав научного коллектива с указанием фамилий, имен, отчеств (при наличии) членов коллектива, их возраста на момент подачи заявки, ученых степеней, должностей и основных мест работы, формы отношений с организацией (трудовой договор, гражданско-правовой договор) в период реализации проекта

1. Крюков Александр Павлович, 67 лет, к.ф.-м.н., МГУ имени М.В.Ломоносова, заведующий лабораторией, трудовой договор.
2. Дубенская Юлия Юрьевна, 39 лет, МГУ имени М.В.Ломоносова, н.с., трудовой договор.
3. Власкина Анна Александровна, 21 год, физический факультет МГУ имени М.В.Ломоносова, бакалавр, трудовой договор.
4. Гресь Елизавета Олеговна, 22 года, физический факультет ИГУ, студент, трудовой договор.

Соответствие профессионального уровня членов научного коллектива задачам проекта

Предложенный проект посвящен, в основном, информационным технологиям, в частности одному из самых его бурно развивающихся разделов – интеллектуальным методам анализа больших данных. Однако, проект имеет также сильно выраженный междисциплинарный характер, а его результаты ориентированы на одну из самых передовых областей научных исследований – астрофизику субатомных частиц, которая занимается изучением самых глубоких основ строения Вселенной. Поэтому в коллективе участвуют специалисты в области как ИТ, так и физики высоких энергий, включая гамма-астрономию.

Руководитель коллектива А.П.Крюков профессионально занимается развитием и применением современных информационных технологий в физике. Им были развиты многочисленные математические методы в области компьютерной алгебры, в том числе для задач физики высоких энергий. Он являлся пионером во внедрении грид технологии в России, выполнил многочисленные исследования в области распределенных вычислений. Имеет интересные физические результаты. А.П.Крюков является членом авторского коллектива коллаборации CMS (ЦЕР, Женева), которая в 2012 году открыла бозон Хиггса (в 2013 году нобелевская премия по физике была присуждена П.Хиггсу и Ф.Энглеру за теоретическое обоснование существования бозона Хиггса). В последние годы он стал активно развивать различные методы машинного обучения и их применения в физике, в частности в гамма-астрономии. Является членом международной коллаборации TAIGA, основной задачей которой является исследование в области физики космических лучей и гамма-астрономии.

Ю.Ю.Дубенская является опытным исследователем, принимала участие во многих проектах по распределенным вычислениям и использованию технологии виртуализации в области высокопроизводительных вычислений. В настоящее

время занимается развитием методов генерации состояний физических систем с помощью генеративно-состязательных нейронных сетей. Результаты докладывались на ряде международных конференций и опубликованы в печати.

А.А.Власкина — бакалавр физического факультета МГУ имени М.В.Ломоносова, которая начинает свой путь в науке. Тем не менее, уже сейчас она добилась хороших результатов в области машинного обучения. Ее курсовая работа посвящена использованию сверточных автоэнкодеров для улучшения отношения сигнал/шум для изображений широких атмосферных ливней, порожденных гамма квантами от галактических и внегалактических источников. В настоящее время по материалам курсовой работы готовится статья для публикации.

Е.О.Гресь — магистрант ИГУ, г.Иркутск. Несмотря на свой возраст уже имеет опыт обработки данных эксперимента TAIGA. В своей бакалаврской работе «Обработка первичных данных ШАЛ Тункинского телескопа TAIGA-IACT» она разработала программу обработки данных телескопа на основе параметров Хилласа, что позволило заметно улучшить качество отбора гамма-событий в эксперименте TAIGA. В настоящее время она занимается задачей классификации событий методом сверточных нейронных сетей. Полученные результаты планируется доложить на международном рабочем совещании «Deep Learning in Computational Physics».

Таким образом, коллектив имеет в своем составе опытных исследователей и студентов, сочетая в себе опыт и знания как в области информационных технологий, включая машинное обучение, так и в области физики высоких энергий, включая гамма-астрономию, методы для которой будут разрабатываться, апробироваться и исследоваться. Все это является залогом успешного выполнения программы исследований и получение заявленных результатов, предусмотренных проектом.

1.8. Планируемый объем финансирования проекта Фондом по годам (указывается в тыс. рублей):

в 2022 г. - 1500 тыс. рублей,

в 2023 г. - 1500 тыс. рублей.

Несоответствие планируемого объема финансирования проекта (в том числе отсутствие информации в соответствующих полях формы) требованиям пункта 10 конкурсной документации является основанием недопуска заявки к конкурсу.

1.9. Научный коллектив по результатам проекта в ходе его реализации предполагает опубликовать в рецензируемых российских и зарубежных научных изданиях не менее

Приводятся данные за весь период выполнения проекта. Уменьшение количества публикаций (в том числе отсутствие информации в соответствующих полях формы) по сравнению с порогом, установленным в пункте 16.2 конкурсной документации, является основанием недопуска заявки к конкурсу.

5 публикаций,

из них

5 в изданиях, индексируемых в базах данных «Сеть науки» (Web of Science Core Collection) или «Скопус» (Scopus).

Информация о научных изданиях, в которых предполагается опубликовать результаты проекта, в том числе следует указать в каких базах индексируются данные издания - «Сеть науки» (Web of Science Core Collection), «Скопус» (Scopus), РИНЦ, иные базы, а также указать тип публикации - статья, обзор, тезисы, монография, иной тип

Результаты проекта планируется опубликовать в ряде международных и российских изданиях, в том числе:

- Journal of Computational Physics,
- Neural Networks,
- Neural Computing and Applications
- Journal of Physics: Conference series
- Программирование
- CEUR Wokrshops

Все указанные издания индексируются системами WoS или Scopus. Точный список изданий будет определен в процессе выполнения проекта.

Иные способы обнародования результатов выполнения проекта

Выступления на ведущих всероссийских и международных конференциях, в том числе следующих регулярных форумах:

- International Workshop on Advanced Computing and Analysis Techniques in Physics Research (ACAT)
- Параллельные вычислительные технологии (ПаВТ), Россия;

-- Distributed Computing and Grid-technologies in Science and Education (GRID), ОИЯИ, Дубна,

-- Суперкомпьютерные дни в России (RSCD), Москва.

Полный список конференций будет определен позднее.

1.10. Число публикаций членов научного коллектива, опубликованных в период с 1 января 2016 года до даты подачи заявки,

53, из них

53 – опубликованы в изданиях, индексируемых в Web of Science Core Collection или в Scopus.

1.11. Планируемое участие научного коллектива в международных коллаборациях (проектах) (при наличии)

В процессе выполнения работ по проекту планируется активное участие в международной коллаборации TAIGA (<https://theory.sinp.msu.ru/doku.php/taiga/about>), в которую входят ведущие российские и европейские научные организации и университеты, в том числе НИИЯФ МГУ; Иркутский государственный университет; Объединенный институт ядерных исследований (ОИЯИ, г.Дубна); Национальный исследовательский ядерный университет "МИФИ"; Институт ядерных исследований РАН; Немецкий электронно-синхротронный центр (Deutsches Elektronen-Synchrotron), DESY, Германия; Туринский университет (Universita' degli Studi di Torino), Италия; Институт технологий (Karlsruher Institut für Technologie), KIT, г. Карлсруэ, Германия; Гамбургский университет (University of Hamburg), УНН, Германия; и ряд других научных организаций и университетов. Основной задачей этой коллаборации является поиск галактических источников гамма-квантов с энергиями выше 20-30 ТэВ; исследования потоков гамма-излучения от известных источников в диапазоне энергий выше 20-30 ТэВ на регистрируемом уровне чувствительности; исследования высокоэнергетической части спектра гамма-излучения наиболее ярких блазаров с целью изучения поглощения гамма-квантов на межгалактическом фоновом излучении (инфракрасном и микроволновом) и поиска возможных нарушений лоренц-инвариантности и аксион-фотонных переходов, а также поиска темной материи во Вселенной. Ключевой идеей развития гамма-обсерватории TAIGA является совместная работа широкоугольных и узкоугольных детекторов Тунка-HiSCORE и TAIGA-IACT. TAIGA – проект класса мегасайенс, находится на территории Тункинского астрофизического центра коллективного пользования Иркутского государственного университета где строится крупнейшая в мире по площади гамма-обсерватория. Руководитель проекта является членом указанной коллаборации.

Руководитель проекта подтверждает, что

- все члены научного коллектива (в том числе руководитель проекта) удовлетворяют пунктам 4, 6, 7, 13 конкурсной документации;
- на весь период реализации проекта руководитель проекта будет состоять в трудовых отношениях с организацией;
- при обнародовании результатов любой научной работы, выполненной в рамках поддержанного Фондом проекта, руководитель проекта и научный коллектив будут указывать на получение финансовой поддержки от Фонда и организацию, а также согласны с опубликованием Фондом аннотации и ожидаемых результатов поддержанного проекта, соответствующих отчетов о выполнении проекта, в том числе в информационно-телекоммуникационной сети «Интернет»;
- помимо гранта Фонда проект не будет иметь других источников финансирования в течение всего периода практической реализации проекта с использованием гранта Фонда;
- проект не является аналогичным по содержанию проекту, одновременно поданному на конкурсы научных фондов и иных организаций;
- проект не содержит сведений, составляющих государственную тайну или относимых к охраняемой в соответствии с законодательством Российской Федерации иной информации ограниченного доступа;
- доля членов научного коллектива в возрасте до 39 лет включительно в общей численности членов научного коллектива будет составлять не менее 50 процентов в течение всего периода практической реализации проекта;
- в установленные сроки будут представляться в Фонд ежегодные отчеты о выполнении проекта и о целевом использовании средств гранта.

Подпись руководителя проекта _____ /А.П. Крюков/