

Сведения о выполненных работах
в период с 27.07.2022 г. по 30.06.2023 г.

по проекту **«Анализ большого массива экспериментальных данных лазерного поляризационного и аэрологического зондирования атмосферы с использованием новейших методов машинного обучения»**,
поддержанному Российским научным фондом

Соглашение № 21-72-10089

Руководитель: Кучинская Олеся Ивановна, канд. физ.-мат. наук

Второй год проекта был посвящен исследованию особенностей применимости искусственных нейронных сетей (ИНС) к задаче восстановления зависимости между метеорологическими параметрами и оптическими характеристиками облаков верхнего яруса (ОВЯ).

Существующие модели, описывающие связи метеорологических параметров атмосферы и характеристик ОВЯ, не позволяют учитывать некоторые атмосферные процессы ввиду упрощенного физического описания среды, либо в рамках этих моделей учет таких зависимостей становится очень нетривиальной задачей. ИНС способны учитывать сложное взаимодействие метеорологических параметров друг с другом, а также восстановить практическую любую зависимость характеристик ОВЯ яруса от этих параметров. В процессе создания прототипа программного продукта (ПП) наибольшую сложность представляло определить архитектуру сети, функцию потерь и способ подачи входных параметров. Каждая из этих проблем влияла на самую главную проблему ИНС – проблему переобучения, когда ИНС перестает обобщать данные и начинает на них настраиваться.

При выборе архитектуры ИНС и одной из возможных моделей машинного обучения мы отталкивались от практической эвристики: количество параметров для настройки модели должно быть меньше набора данных, используемых для обучения. При выполнении данного условия ИНС начинает обобщать данные, а не запоминать их, что позволяет ограничить набор потенциальных архитектур, которые могут быть использованы для обучения. С этой целью произведен предварительный анализ данных лидарных измерений матрицы обратного рассеяния света (МОРС) ОВЯ за период 2016–2020 гг. Датасет был обобщен в базу данных, которая содержит информацию о 1177 сериях измерений. Из них в 312 случаях были обнаружены ОВЯ. Кроме того, еще в 172 сериях также зарегистрированы ОВЯ, но лидарный сигнал от них был недостаточен для статистической обработки. Эти данные были использованы для проверки прогнозирования ИНС высот формирования ОВЯ в зависимости от метеорологических условий. Результаты лидарных измерений сопоставлены с данными реанализа ERA5 в каждой точке высотного профиля от 0 до 15 км с шагом в 150 метров. Для каждого сеанса с учетом данных ERA5 сформированы высотные профили следующих метеорологических параметров: относительная и абсолютная влажность, скорость и направление ветра, температура.

Далее, была создана нейросеть для решения следующих задач атмосферной оптики:

- 1) определить вероятность наблюдения ОВЯ в зависимости от метеорологических параметров (задача классификации);
- 2) осуществить предварительную оценку высоты наблюдения и границ ОВЯ в зависимости от метеорологических параметров (задача регрессии);
- 3) оценить значения элементов МОРС ОВЯ, используя метеорологические параметры (задача регрессии).

Несмотря на существенное количество лидарных экспериментов, количество серий измерений 2016–2020 гг., в течение которых регистрировались ОВЯ и которые пригодны для обучения нейросети, составило всего 312. В течение года, в экспериментальный массив были добавлены результаты измерений 2021–2023 гг. (50 серий), а также дополнительные данные 2009–2016 гг. (809 серий). Проблема статистической малости массива экспериментальных данных была решена программно, что приемлемо на этапе разработки прототипа ПП. Следующим шагом необходимо больше данных для уточнения результата и добавление дополнительных сведений об окружающей среде в области проведения эксперимента.

Анализ полученных результатов показал, что имеется зависимость от метеорологических характеристик у следующих величин: высота формирования ОВЯ, элементы m_{22} и m_{44} его МОРС. Поиск таких зависимостей для m_{33} требует дальнейшего изучения и наращивания статистики. Элемент m_{11} всегда равен 1 (поскольку на него нормируются все элементы МОРС). Остальные величины определяются с достаточно большими погрешностями, и дать однозначный ответ об их связях с метеопараметрами на текущий момент невозможно.

Почти во всех случаях метод случайного леса дал более качественный результат, чем простой многослойный перцептрон. Наиболее вероятно, это вызвано двумя факторами: предобработка входных данных и небольшое количество экспериментальных данных пригодных для обучения. В качестве дальнейших шагов необходимо рассмотреть преобразования, приводящие гистограммы распределение входных данных к симметричному виду и более похожему на нормальное распределение.

Размер данных также повлиял на работу ИНС. При усложнении ИНС она способна аппроксимировать более сложные функции, но при этом необходимо использовать больший объем данных. Использование небольшого объема данных приводило к переобучению сети и, соответственно, к ухудшению получаемых результатов. Для улучшения результатов, а также для установления дополнительных возможных зависимостей между метеорологическими параметрами и характеристиками ОВЯ мы планируем провести исследования в предварительной обработке метеорологических величин, а именно: учет динамики изменения с течением времени, градиент в области эксперимента, учет антропогенного фактора. Также необходимо наращивание объема массива экспериментальных данных (выполнение новых экспериментов по лазерному зондированию атмосферы).

Как показал опыт работы с прототипом ПП, для корректной работы требуются данные серий лидарных измерений, в течение которых сигнал от ОВЯ характеризовался отношением сигнал-шум, превышающим единицу. Поскольку зондирование ОВЯ проводится круглосуточно, в дневное время, вследствие повышенной яркости неба, уровень помехи на фотодетекторе существенно возрастает. Это в значительной степени увеличивает погрешность оценки лидарного сигнала от ОВЯ и, тем самым, приводит погрешностям определения элементов МОРС. Закупка в рамках гранта оптических элементов позволила частично решить проблему увеличения отношения сигнал/шум за счет уменьшения потерь в передающей и приемной системах лидара.

Результаты работы опубликованы в 6 статьях, 4 из которых входят в перечень WOS и Scopus, при этом одна статья – Q1 (<https://www.mdpi.com/2072-4292/15/1/109>), вторая – Q2 (<https://www.mdpi.com/2073-4433/13/8/1225>). Сделано пять научных докладов и два выступления запланированы в июне на международных симпозиумах МСАРД-2023 (г. Санкт-Петербург) и ОАиО (г. Москва). Руководитель проекта представила материалы гранта в конкурсной программе на соискание премии «Женщины в науке и технологиях KOLBA» и стала победителем в номинации «Физика» (<https://news.tsu.ru/news/fizik-tgu-olesya-kuchinskaya-poluchila-natsionalnuyu-premiyu-dlya-zhenshchin-uchenykh/>). Была приглашена в качестве спикера в дискуссионной сессии «Кибербезопасность и метаданные» (<https://perspectum.info/forum-i-premiya-kolba-2023>).

Результаты проекта использованы для написания диссертации на соискание ученой степени канд. физ.-мат. наук. «Оптические свойства облаков верхнего яруса естественного и антропогенного происхождения, содержащих ориентированные кристаллы льда, по данным поляризационного лазерного зондирования», Брюханов И.Д., (http://kirensky.ru/ru/councils/dis/dissertacii-prinyatye-k-zaschite-1/diss_bryuhanov_id).