

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования

«ПЕРМСКИЙ ГОСУДАРСТВЕННЫЙ НАЦИОНАЛЬНЫЙ
ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ»

ЦИФРОВАЯ ГЕОГРАФИЯ

Материалы Всероссийской научно-практической конференции
с международным участием

16–18 сентября 2020 г., г. Пермь

Том I

Цифровые и геоинформационные технологии в изучении природных
процессов, экологии, природопользовании и гидрометеорологии



Пермь 2020

ДЕТЕКТИРОВАНИЕ ГРОЗОВЫХ ЯЧЕЕК ПО ДАННЫМ СПУТНИКОВОГО ЗОНДИРОВАНИЯ

В работе рассмотрена возможность использования данных микроволновой интегрированной восстанавливающей системы (MiRS) расширенных дополнительными расчетными параметрами атмосферы и алгоритма машинного обучения (XGBoost) для детектирования грозовых ячеек.

Ключевые слова: Западная Сибирь, машинное обучение, дистанционное зондирование.

V.V. Chursin, I.V. Kuzhevskaya, skriptym@mail.ru

National Research Tomsk State University, Tomsk, Russia

DETECTING OF THUNDERSTORM CENTERS USING BY SATELLITE REMOTE SENSING

The paper considers the possibility of using MiRS data extended by additional computed atmospheric parameters and the machine learning algorithm (XGBoost) for the detection of thunderstorm cells.

Keywords: Western Siberia, machine learning, remote sensing.

В настоящее время отмечается растущая потребность хозяйственных подразделений в информации о различных метеоэлементах, особенно об опасных и неблагоприятных погодных явлениях (ОЯ). Интенсивное развитие в строительстве инженерных сооружений, освоение природных ресурсов, модернизации транспортных сетей и рост сельскохозяйственной активности требуют все более тщательного подхода при прогнозировании и диагнозе ОЯ, в частности зон возможного развития грозовых ячеек. Согласно статистике, вероятность поражения человека молнией в течение жизни составляет 1:3000 (для территории США) [2]. Хотя вероятность и кажется достаточно малой, однако, например, на территории США в период между 1995 и 2008 гг. от удара молний погибли 648 человек. Точное диагностирование грозовых ячеек в свою очередь невозможно без привлечения дистанционных методов исследования и современных методик и физико-математических алгоритмов [3].

Одним из наиболее востребованных дистанционных методов исследования атмосферы является спутниковое зондирование Земли. В настоящей работе были использованы данные микроволновой интегрированной восстанавливающей системы (MiRS) с 13 мая по 27 июня 2019 г., предоставленные Сибирским центром Научно-исследовательского центра космиче-

ской гидрометеорологии «Планета» [1].

The Microwave Integrated Retrieval System (MiRS) – это итеративный физический одномерный алгоритм восстановления (1DVAR), принцип которого заключается в минимизации функции потерь из двух источников. Эта функция состоит из отклонения моделируемых излучений от фактических измерений и отклонения найденных параметров от их соответствующих фонов. Процесс восстановления выполняется в ограниченном пространстве и ожидает несколько степеней свободы, чтобы обеспечить баланс между стабильностью матрицы и оптимальным извлечением полезной информации. Система MiRS была разработана, чтобы быть гибким инструментом восстановления / ассимиляции и поэтому подходит для применений в микроволновом зондировании, а также в других областях спектра, хотя в основном применяется только к приборам миллиметровых частот. Эти функции в сочетании с усовершенствованной моделью радиационного переноса (CRTM), используемой в качестве прямого оператора, позволяют MiRS быть передовым алгоритмом, легко применимым к текущим и будущим датчикам, как зондирующим, так и визуальным.

Метеорологические спутники способны обеспечивать наблюдения с помощью дистанционного зондирования внутри и вокруг зон развития грозовых ячеек с вы-

соким горизонтальным, вертикальным и/или временным разрешением. Особый интерес для наблюдений таких ячеек и их численного прогнозирования представляют инфракрасные и микроволновые приборы. Полярно-орбитальный спутник обеспечивает глобальные измерения излучения в микроволновой и инфракрасной областях до двух раз за сутки. Инфракрасный радиометр чрезвычайно полезен для измерения отражаемости в многоканальном режиме для профилирования температуры воздуха и водяного пара с высоким вертикальным разрешением [4, 5]. Однако инфракрасные каналы не могут проникать сквозь облака, за исключением оптически тонких облаков (например, перистых), в то время как грозовые ячейки представляют из себя в основном плотные кучево-дождевые облака. Напротив, микроволновые приборы не обеспечены таким количеством каналов как инфракрасные радиометры, но могут обеспечить измерения отражаемости для профилирования атмосферной температуры и водяного пара практически во всех погодных условиях, за исключением сильных осадков [6]. Следовательно, микроволновые приборы являются незаменимыми для наблюдения за зонами развития гроз, занятыми облачностью, а инфракрасные радиометры важны для наблюдения за окружающими безоблачными зонами.

Для разработки алгоритма детектирования грозовых ячеек необходимо два массива данных. Первоначально был собран массив, включающий в себя фактические данные об метеорологических явлениях со 134 станций Западной Сибири за 8 метеорологических сроков в период с 13 мая по 27 июня 2019 г. (ВНИИГМИ-МЦД [2]) с присвоением географических координат метеорологических станций на которых наблюдались погодные явления. В итоге мы получаем массив, содержащий в себе координаты станции, дата, время и шифр явления. Из него мы отбираем только «грозы» и «грозы в окрестности» (шифр 80, 81) и присваиваем этим строкам идентификационный номер (1), всем остальным присваиваем (0).

По значениям приземного давления, температуры воздуха и удельной влажности рассчитанной системой MiRS, дополнительно нами были рассчитаны следующие величины:

- Относительная влажность на 1000гПа; 850гПа; 700гПа и 500гПа.
- Температура точки росы на

1000гПа; 850гПа; 700гПа и 500гПа.

- Дефицит точки росы на 1000гПа; 850гПа; 700гПа и 500гПа.
- Vertical Totals.
- Cross Totals.
- Total Totals.
- K index.
- Нижняя граница конвективной облачности (Среднее из 3 методов: Селезнева, Феррель, Ипполитова).
- Сумма дефицитов точки росы по 4 уровням.
- Разность температуры на 850 гПа и частицы поднятой до этого уровня по влажной адиабате с 1000гПа.
- Разность температуры на 700 гПа и частицы поднятой до этого уровня по влажной адиабате с 1000гПа.
- Разность температуры на 500 гПа и частицы поднятой до этого уровня по влажной адиабате с 1000гПа
- Среднее значение от 3 предыдущих переменных.

Затем формировался второй массив данных спутникового зондирования путем осреднения значений над каждой метеорологической станцией с радиусом 0,2° (десятичная система) за каждые 3 часа с космических аппаратов MetopB, Suomi NPP, NOAA-19. В случае перекрытия зон зондирования несколькими спутниками, отбиралось их среднее значение.

После создания вышеописанных массивов было произведено их слияние по дате, времени и географическим координатам, что позволило нам получить полноценный массив, содержащий в себе идентификационный номер наличия/отсутствия грозы и соответствующие им физические характеристики атмосферы.

Следующим шагом идет выбор алгоритмов и его обучение для возможности детектирования гроз. Было решено использовать градиентный бустинг, а именно «XGBoost classifier». В основе XGBoost лежит алгоритм градиентного бустинга деревьев решений. Градиентный бустинг – это техника машинного обучения для задач классификации и регрессии, которая строит модель предсказания в форме ансамбля слабых предсказывающих моделей, обычно деревьев решений. Обучение ансамбля проводится последовательно. На каждой итерации вычисляются отклонения предсказаний уже обученного ансамбля на обучающей выборке. Следующая модель, которая будет добавлена

в ансамбль, будет предсказывать эти отклонения. Таким образом, добавив предсказания нового дерева к предсказаниям обученного ансамбля, мы можем уменьшить среднее отклонение модели, которое является «таргетом» оптимизационной задачи. Новые деревья добавляются в ансамбль до тех пор, пока ошибка уменьшается, либо пока не выполняется одно из правил «ранней остановки» [7].

Для обучения и тестов полученных моделей диагноза грозовых ячеек было сгенерировано три набора данных для тестов.

Генерация первой выборки осуществляется путем выполнения следующей последовательности:

1. Из итогового массива выбираются строки, содержащие идентификационный номер наличия грозы (1).
2. По выбранным строкам рассчитывается среднее значение каждого параметра атмосферы и стандартное отклонение.
3. Рассчитывается нижний и верхний пороги для последующей фильтрации (среднее $\pm$ 2 стандартных отклонения).
4. Производим фильтрацию каждой переменной в соответствии с расчётным нижним и верхним пределом на оставшихся после первого пункта строках (без гроз).
5. Получаем 7229 строк, подходящих под параметры, но без фиксирования грозы и 506 строк, когда гроза была зафиксирована на станции.

6. Поскольку выбора несбалансированная, искусственно увеличиваем случаи с грозой до 7000 путем дублирования.

7. Производим слияние двух массивов (гроза/без грозы) и перемешиваем в случайном порядке.

8. Сформированный массив подаем для обучения и тестирования точности (делим выборку на обучающую и тестовую 70/30 %).

Вторая выборка составлялась путем выбора всех случаев без грозы и всех случаев с грозой без модификаций.

Третья выборка составлялась путем выявления эксклюзивных значений. (поскольку в один и тот же срок может быть записано несколько явлений, целесообразно выбрать только эксклюзивные: 1 срок – 1 явление).

Обучение производилось неклассическим способом. Производим перебор глубины дерева, количества слоев, и градиента, но точность оцениваем не по тестовой выборке из первого сгенерированного массива (30 %), а по второй (без модификаций) выборке.

Точность обнаружения грозовых ячеек достигается достаточно быстро, но количество ложных срабатываний велико на ранних этапах, поэтому целесообразно строить глубокую и сложную модель.

Из полученных путем перебора моделей выбрано три с наилучшими показателями точности и разной структурой (параметры моделей приведены в табл. 1).

Таблица 1. Оценки точности полученных моделей на сгенерированных выборках

XGBoost	Метрика	Модель 1 Глубина дерева:6 Количество деревьев: 12000	Модель 2 Глубина дерева:7 Количество деревьев: 10000	Модель 3 Глубина дерева:8 Количество деревьев: 10000
Обучение	Accuracy	0,980	0,980	0,980
	Precision	0,960	0,960	0,960
	recall	1,000	1,000	1,000
	F-мера	0,980	0,980	0,980
Тестовая	Accuracy	0,948	0,953	0,951
	Precision	0,902	0,910	0,907
	recall	1,000	1,000	1,000
	F-мера	0,948	0,953	0,951

XGBoost	Метрика	Модель 1 Глубина дерева:6 Количество деревьев: 12000	Модель 2 Глубина дерева:7 Количество деревьев: 10000	Модель 3 Глубина дерева:8 Количество деревьев: 10000
Выборка 2	Accuracy	0,992	0,991	0,992
	Precision	0,717	0,699	0,724
	recall	1,000	1,000	1,000
	F-мера	0,835	0,823	0,840
Выборка 3	Accuracy	0,995	0,996	0,995
	Precision	0,976	0,981	0,976
	recall	1,000	1,000	1,000
	F-мера	0,988	0,990	0,988

Как видно из таблицы, рассчитываемые параметры атмосферы и предложенная методика обучения моделей, применимо к MiRS, дают достаточно высокие показатели оправдываемости диагноза грозовых ячеек, что, в свою очередь, при

значительном пространственном и временном разрешении, может положительно сказаться на обнаружении гроз в регионах, которые не имеют аэрологического или радиолокационного обеспечения.

Библиографический список

1. Сибирский центр Научно-исследовательского центра космической гидрометеорологии «Планета» [Электронный ресурс] URL: <https://rcpod.ru/cgi-bin/prod.pl?Srcs=1366!768&r=0.13331125515261566>.
2. Специализированные массивы данных ВНИИГМИ-МЦД. [Электронный ресурс] URL: <https://http://meteo.ru/data>.
3. Хандожко Л.А. Экономическая метеорология. Гидрометеиздат, СПб. 2005, 489 с.
4. Chen H., Zhang D.-L. On the rapid intensification of Hurricane Wilma 2005. Part II: Convective bursts and the upper-level warmcore // J. Atmos. Sci., 2013. 70. P.146-162.
5. Chen T., Singh S., Taskar B., and Guestrin C. Efficient second-order gradient boosting for conditional random fields. // In Proceeding of 18th Artificial Intelligence and Statistics Conference (AISTATS'15), volume 1, 2015.
6. Flash Facts About Lightning // National Geographic News, June 24, 2005.
7. Hilton, F., et al. Hyperspectral Earth observation from IASI: Five years of accomplishments // Bull. Am. Meteorol. Soc., 2012. 93(3). P. 347-370.
8. Weng F., Zhao L., Ferraro R.R., Poe G., Li X., and Grody N.C. Advanced microwave sounding unit cloud and precipitation algorithms, Radio Sci., 2003. 38 (4). 8068.