

ИСПОЛЬЗОВАНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ СОЗДАНИЯ КАТАЛОГА СОЛНЕЧНЫХ ВСПЫШЕК ПО НАБЛЮДЕНИЯМ НА СИБИРСКОМ РАДИОГЕЛИОГРАФЕ

Шамсутдинова Ю.Н., Рожкова Д.В., Кашапова Л.К., Губин А.В.
Институт солнечно-земной физики СО РАН, Иркутск, Россия

USING MACHINE LEARNING METHODS TO CREATE A CATALOG OF SOLAR FLARES BASED ON OBSERVATIONS AT THE SIBERIAN RADIOHELIOGRAPH

Shamsutdinova J.N., Rozhkova D.V., Kashapova L.K., Gubin A.V.
Institute of Solar-Terrestrial Physics SB RAS, Irkutsk, Russia

We present the results of designing a catalog of solar flares using machine learning methods based on the Siberian Radioheliograph (SRH). A technique for selecting candidate events has been proposed and tested, allowing determining the onset, maximum, and end times of events by analyzing their time profiles, presented as a table of functions. Since the SRH is obtaining multi-wave observations, we introduced a criterion for selecting events based on the simultaneous response at several frequencies, which allowed us to automatically select events. Currently, the technique focuses on selecting broadband events in order to avoid the appearance of narrow-band artifacts. During the period from May to June 2023, a total of about 729 events with different power classes were identified. During this same period in 2024, 829 solar flares were detected. Verification of selected events, such as simple and complex solar flares, was done by applying the Support Vector Machine (SVM) method using a trained and tested model. The algorithm was applied in test mode to identify solar flares observed in the 3–24 GHz range during the summers of 2023 and 2024. Based on the selected events, the possibility of using the SVM method to recognize complex and simple time profiles of microwave flare emission was tested based on correlation plot profiles observed by SRH. Parameters were selected to achieve the best possible recognition results for both simple and complex time profiles of solar flares in microwaves.

<https://doi.org/10.31725/0552-5829-2024-347-350>

Объем информации, получаемой при наблюдении Солнца, резко возрастает в связи с появлением новых приборов с высоким временным разрешением и чувствительностью, охватывающих широкий диапазон электромагнитного спектра. Это привело к необходимости разработки и усовершенствования автоматизированных и надежных методов обнаружения событий и их каталогизации в больших потоках данных. Используемый в работе подход основан на анализе производной временного профиля, который представлен в виде непрерывной числовой функции в сочетании с методами машинного обучения, которые в последнее время доказали свою эффективность.

Поскольку эволюция солнечной вспышки описывается как внезапное увеличение потока, за которым следует непрерывный спад, мы можем использовать производную функцию для определения начала, максимума и окончания события. Прежде всего, для каждой спектральной полосы или диапазона длин волн вычисляется стандартное отклонение (сигма) производной временного профиля в интервале времени, где отсутствуют события. Если значение производной временного профиля находится вне интервала, определяемого как три сигмы, то считается, что оно соответствует событию. Начало события определяется как момент времени, где значения производной от близких к нулю становятся положительными, что указывает на резкое возрастание временного профиля, а конец события – момент, где производная знак с отрицательного на близкий к нулю. Точный момент максимума определяется как максимальное значение временного профиля между найденными моментами начала и окончания события.

Разработанную методику мы применили к наблюдаемым данным Сибирского Радиогелиографа (СРГ). Инструмент представляет собой интерферометр, состоящий из трех независимых антенных решеток, диапазоны частот которых составляют 3–6, 6–12, 12–24 ГГц [1]. Для обнаружения солнечных вспышек в наблюдениях СРГ мы использовали один из видов данных – корреляционные кривые, представляющие собой сумму комплексных ковариаций, вычисляемых для различных пар антенн. Благодаря чувствительности этих корреляционных кривых можно выявлять и исследовать слабые всплески микроволнового излучения [2]. Поэтому именно эти временные профили лучше всего подходят для целей нашего исследования. Поскольку СРГ является многоволновым инструментом, то разработанный алгоритм был дополнен возможностью автоматически отбирать события в зависимости от одновременного отклика на нескольких частотах. В большинстве случаев узкополосный всплеск может быть артефактом. Поэтому, для того чтобы обнаруженный всплеск считался вспышкой, необходимо, чтобы наблюдался отклик как минимум на 15 частотах или не менее 5 частот в двух спектральных диапазонах. Так как моменты максимумов всплесков, относящихся к одному и тому же событию, могут не совпадать на разных частотах из-за особенностей наблюдений на СРГ или физических факторов, то мы предполагаем, что всплески с разной частотой относятся к одному и тому же событию с интервалом в 200 секунд. Тестирование методики происходило на наблюдениях за 13 июля 2023 года. Результат показан на рисунке 1. Можно видеть, что 18 событий удовлетворяют строгим критериям отбора, причем некоторые из них были слабыми и почти неразличимы на временном профиле 1–8 А, полученным КА GOES. В общей сложности за период с мая по июнь 2023 года мы выявили около 729 событий различного класса мощности, а в аналогичный период в 2024 году произошло около 829 вспышек на Солнце. Полученные результаты позволили обеспечить объем данных, достаточный для проведения

тестирования возможности использования методов машинного обучения для классификации временной структуры излучения солнечных вспышек в микроволновом диапазоне. Предварительные результаты представлены на сайте: stmw.iszf.irk.ru.

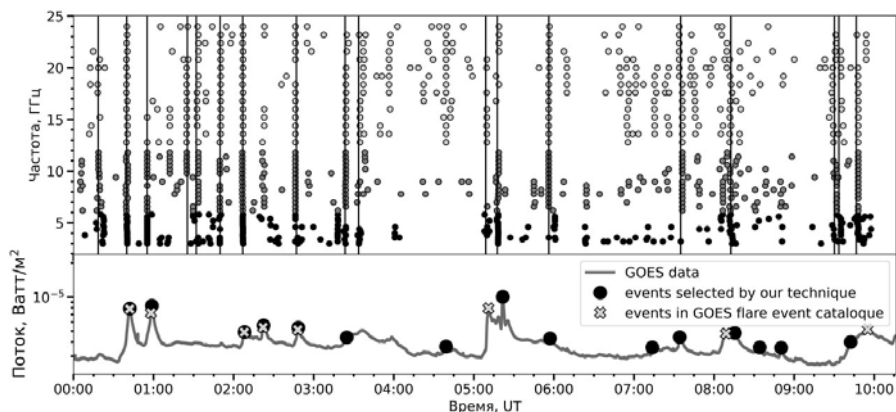


Рис. 1. Пример обнаружения вспышек в диапазоне 3–24 ГГц для наблюдений 13 июля 2024 года. *Верхняя панель* показывает диаграмму микроволновых всплесков, обнаруженных по данным SRG. Круги, черные окружности и серые окружности показывают диапазоны частот 3–6 ГГц, 6–12 ГГц и 12–24 ГГц соответственно. Вертикальные линии показывают отобранные события. *Нижняя панель:* временной профиль рентгеновского излучения по наблюдениям КА GOES (серая линия). Крестики показывают события, указанные в каталоге КА GOES, кружки – события, отобранные с использованием оригинальной методики.

Для классификации отобранных событий был протестирован метод опорных векторов (Support Vector Machine, SVM), используя временные профили. SVM – это алгоритм машинного обучения, применяемый для задач классификации и получения регрессии. Основная цель алгоритма – найти оптимальную гиперплоскость в высокоразмерном пространстве, которая разделяет точки данных на различные классы в пространстве признаков. Когда количество входных признаков больше трех, SVM использует ядра для преобразования гиперплоскости в нелинейную, такие как линейная, полиномиальная и радиальная базисная функция ядра (RFB).

Набор данных для обучения и тестирования состоит из временных профилей микроволнового излучения на частоте 9.4 ГГц. Выбор частоты был обусловлен тем, что на этой частоте происходит наибольшее количество солнечных вспышек, которые формируются в области оптически тонкого источника и, следовательно, дополнительные эффекты, связанные с нагревом и движением вещества, не должны наблюдаться. Чтобы использовать эти данные в машинном обучении, их нужно было привести к общему виду и подготовить для работы. Мы нормировали каждый временной профиль на его максимальную интенсивность. Момент максимума использовался как нулевая точка на временной шкале. Таким образом, время в фазе роста имеет отрицательные значения, а время в фазе спада – положи-

тельные. Для того чтобы сопоставить события разной длительности в рамках единой временной шкалы, время было приведено к единой единице измерения. Эта единица была определена как промежуток между моментами времени, когда интенсивность события была равна половине от максимума для временного профиля каждого события [3]. Пример нормализации данных по времени представлен на рисунке 2(а).

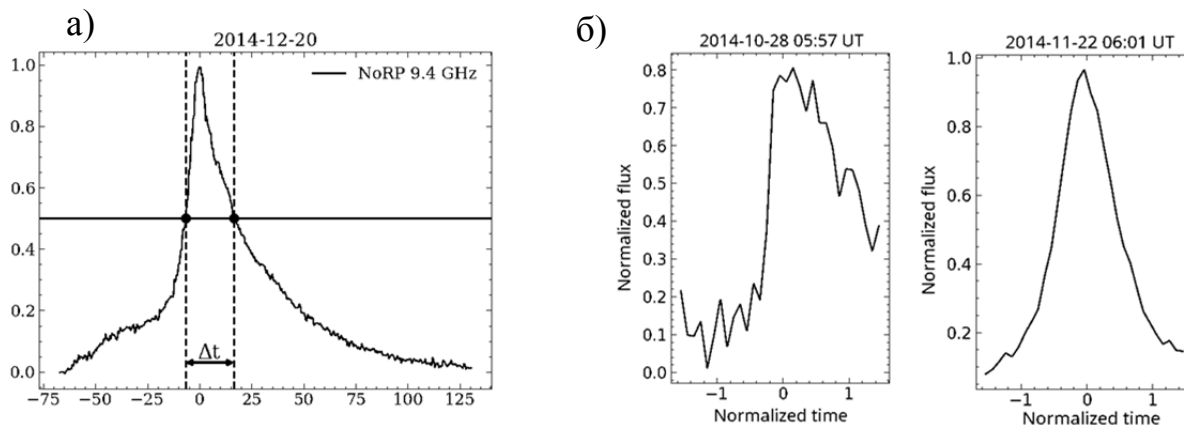


Рис. 2. (а) Пример предварительной обработки временного профиля и поиска временного шага для нормирования. (б) Пример набора входных данных после применения первичной обработки. Левая панель: временные профили со сложной структурой (класс = 0). Правая панель: временные профили с классической структурой (класс = 1).

Каждый массив содержит 200 элементов, что соответствует количеству признаков. Для обучения и тестирования мы использовали набор данных, состоящий из 200 событий с июня по август 2023 года. Примеры профилей представлены на рисунке 2(б). 84 профиля имеют сложную структуру, которая может содержать несколько пиков и/или шум, доминирующий над сигналом (класс 0). Остальные 116 профилей имеют классическую структуру (класс 1). Результат тестирования показал, что наиболее точную информацию получило ядро RFB с параметрами γ и C , равными 1. Однако полученные результаты необходимо проверить на большем наборе данных в широком диапазоне частот.

Исследование выполнено за счет гранта Российского научного фонда № 24-22-00315 (<https://rscf.ru/project/24-22-00315/>).

Литература

1. Алтынцев А.Т. и др. // Солнечно-земная физика, 2020, т. 6, в. 2, с. 37-50.
2. Лесовой С.В. и В.С. Кобец // Солнечно-земная физика, 2017, т. 3, в. 1, с. 17-21.
3. Davenport J.R.A., et al. // Astrophysical Journal, 2014, v. 797, p. 122.