

ПРИМЕНЕНИЕ МАШИННОГО ОБУЧЕНИЯ ДЛЯ РАСПОЗНАВАНИЯ СОЛНЕЧНЫХ ВСПЫШЕК В МИКРОВОЛНОВОМ ДИАПАЗОНЕ

Ю.Н. Шамсутдинова, Л.К. Кашапова

Институт солнечно земной физики Сибирского отделения Российской академии наук,
Иркутск, Россия, yulia@iszf.irk.ru

THE RECOGNITION OF SOLAR FLARES IN MICROWAVE RANGE USING MACHINE LEARNING

Yu.N. Shamsutdinova, L.K. Kashapova

Institute of Solar-Terrestrial Physics SB RAS, Irkutsk, Russia, yulia@iszf.irk.ru

Аннотация. Мы представляем результаты тестирования метода опорных векторов для распознавания солнечных вспышек в микроволновом диапазоне. Для обучения использовались данные из каталога событий спектрополяриметра Nobeyama (NoRP). В качестве входных данных были использованы временные профили на частоте 9.4 ГГц за 2001, 2012, 2013, 2014 гг. Набор данных состоял из 100 событий и включал в себя временные профили вспышек с простой и сложной структурой. Максимальная точность распознавания получена при функции радиального ядра и значении параметра гамма равном единице.

Ключевые слова: машинное обучение, солнечные вспышки, микроволновый диапазон.

Abstract. We present the results of testing a method for recognizing solar flares in the microwave range using machine learning. We used Nobeyama spectropolarimeter (NoRP) catalog for training. Temporal profiles at the frequency of 9.4 GHz for 2001, 2012, 2013, 2014 were used as the dataset. The dataset consisted of 100 events and included temporal profiles with simple and complex structure. Maximum recognition accuracy is obtained with radial kernel function and gamma parameter equal to one.

Keywords: Machine Learning, solar flare, microwave range, temporal profile.

ВВЕДЕНИЕ

В настоящее время в физике Солнца проводятся исследования, направленные на использование методов машинного обучения для прогнозирования возникновения солнечных вспышек [Nishizuka et al., 2017]. В то же время выявление событий среди потока информации, поступающей с новых инструментов, становится все более и более актуальной. В настоящее время развиваются методы распознавания различных типов радиовсплесков на динамических спектрах [Xu et al., 2019]. Однако для выявления и идентификация солнечных вспышек в микроволновом диапазоне, как и для большинства одномерных наблюдений в других спектральных диапазонах подбор и тестирование использования методов машинного обучения не проводилось. Мы представляем первые результаты тестирования использования метод опорных векторов (SVM) для выявления солнечных вспышек с классической формой временного профиля. Считается, что рассматриваемый метод эффективен для данных большого объема данных. Для создания модели была использована библиотека scikit-learn для языка Python (<https://scikit-learn.org>).

НАБОР ДАННЫХ

Тестирование было проведено с использованием временных профилей микроволнового излучения из базы данных событий, наблюдавшихся спектрополяриметром Nobeyama (NoRP) [Torii et al., 1979]. NoRP наблюдает за Солнцем в микроволновом диапазоне на частотах 1, 2, 4, 9.4, 17 и 35 ГГц с временным разрешением до 0.1 секунды. Инструмент измеряет интенсивность от всего Солнца (Солнце как звезда) и круговую поляризацию V. В качестве данных для обучающей и тестовой выборки были использованы события из каталога NoRP за 2001, 2012–2014 гг.

[<https://solar.nro.nao.ac.jp/norp/html/event/>]. Данные года были периодами высокой солнечной активности. Мы использовали только потоки интенсивности, который изначально измеряются в солнечных единицах потока (СЕП). Пример временного профиля представлен на рис. 1 (верхняя панель).

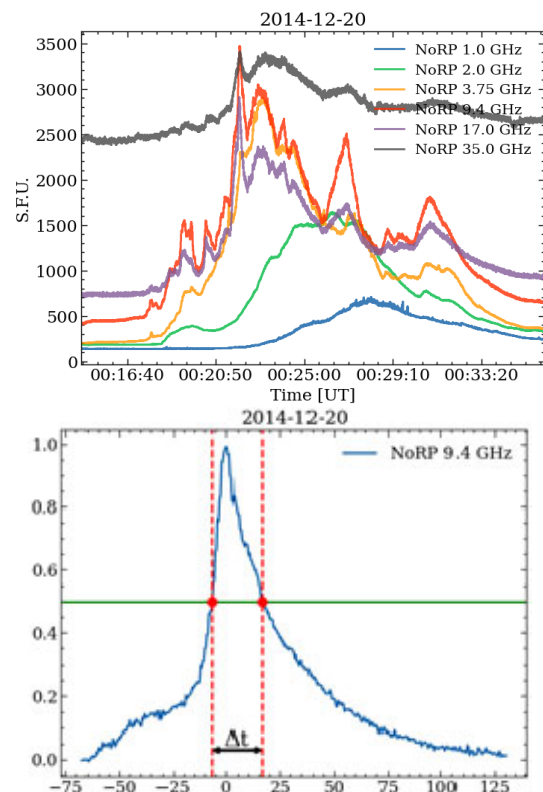


Рис. 1. Верхняя панель: пример временных профилей интенсивности (I) в диапазоне частот 1–35 ГГц по данным NoRP. Нижняя панель: пример предварительной обработки временного профиля. Поиск временного шага для нормирования.

Для распознавания была выбрана временные профили на частоте 9.4 ГГц. Излучение большинства солнечных вспышек на этой частоте формируется в области оптически тонкого источника, что позволяет наблюдать микроволновое излучение без влияния различных эффектов, искажающих временной профиль. Кроме того, эта же частота присутствует в диапазоне частот, на которых наблюдает Сибирский Радиогелиограф (СРГ). Таким образом, результаты, полученные при обучении, могут быть применены к данным СРГ на этой частоте. Так как солнечные вспышки представляют собой события различной длительности и мощности, то они должны быть предварительно обработаны для ввода их в модель машинного обучения.

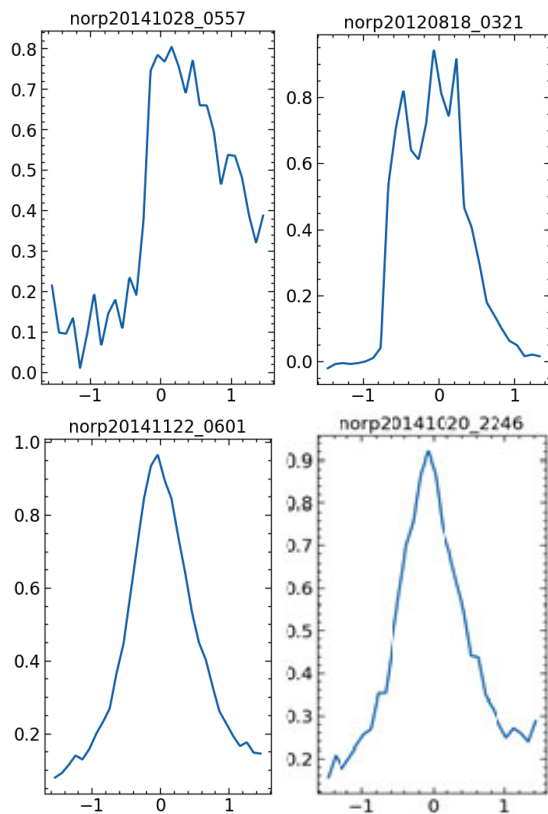


Рис. 2. Пример набора входных данных после применения первичной обработки. Левая панель: временные профили со сложной структурой (класс=0). Правая панель: временные профили с тонкой структурой (класс=1).

Предварительная обработка состояла в том, что в начале находился максимум потока на временном профиле, и он принимался за нулевой отсчет по времени. Таким образом время фазы роста имеет отрицательные значения, а время фазы спада — положительные. Для получения одинаковой шкалы времени для событий с разной продолжительностью было проведено нормирование времени на величину, равную времени между моментами, когда значения потока были равны половине максимума. Описанная процедура применяется для реконструкции среднего профиля вспышки в работе [Davenport et al., 2014]. Каждый массив, содержащий информацию о событии из набора данных, имеет размер 35 элементов. Значение каждого элемента является признаком, который используется как входной параметр для модели. Для обучения и тестирования

мы использовали выборку данных, включающую 100 массивов, из которых 43 профиля имеют сложную структуру (класс 0) и могут оказаться артефактами (например, шумом), а 57 профилей имеют классическую структуру (класс 1). Примеры входных данных на рис. 2. Чтобы повысить точность распознавания, мы дополнительно удаляем неполные события, где временной профиль был прописан не полностью по инструментальным причинам, и применяем кросс-валидацию. Кросс-валидация заключается в разделении имеющегося набора данных на несколько частей, которые используются для обучения модели и тестирования ее качества в произвольной комбинации друг с другом. Этот процесс повторяется несколько раз в различных комбинациях. Такой подход позволяет более эффективно обучить модель на небольшой выборке данных

МЕТОД ОПОРНЫХ ВЕКТОРОВ

Метод опорных векторов (Support Vector Machine, SVM) — алгоритм машинного обучения, который использует модели обучения с учителем для решения задач классификации и регрессии. SVM строит гиперплоскость в многомерном пространстве для разделения различных классов. Следовательно, основная задача SVM заключается в нахождении максимальной граничной плоскости, которая наилучшим образом разделит набор данных на классы. Опорными векторами (support vectors) называются точки данных, которые находятся ближе всего к гиперплоскости. Эти точки определяют разделительную линию и границы (margin). Алгоритм SVM реализуется с помощью ядра, что позволяет построить более точный классификатор. Ядро преобразует пространство входных данных в требуемую форму путем преобразования плоскости низкой размерности в плоскость более высокой размерности. Существуют линейное ядро, полином ядро, радиальная базисная функция (RFB). RFB использует гиперпараметр γ , который определяет влияние опорного вектора на данные. Чем больше γ , тем ближе другие примеры должны быть затронуты. C — это гиперпараметр регуляризации правильности классификации. В нашей работе мы использовали метод GridSearch для определения оптимальных гиперпараметров модели с различными функциями ядер. Мы оптимизировали модель с помощью GridSearch, выбрав гиперпараметры на основе точности, достигнутой в тестовом наборе. Оптимальные параметры для моделей представлены в табл. 1.

SVM широко используется для решения задач бинарной классификации, а также в случаях, когда данные не являются линейно разделимыми. Использование кросс-корреляции предполагает, что временные ряды являются стационарными и имеют линейные отношения, что в нашем случае это не так. Поэтому кросс-корреляция не всегда может быть применена для задачи распознавания.

РЕЗУЛЬТАТ КЛАССИФИКАЦИИ ВРЕМЕННЫХ ПРОФИЛЕЙ И ОЦЕНКА

Для оценки модели классификации мы использовали такие метрики как точность (precision), пол-

нота (recall) и меру F1 как в работе [Wang et al.]. Precision оценивает точность предсказания для истинного положительного класса. Recall показывает процент правильных положительных ответов по отношению к общему количеству фактических положительных результатов. Оценка F1 — это средневзвешенное гармоническое значение precision и recall. Чем ближе к 1, тем лучше модель. TPR означает, что классификатор верно отнес профиль к нужному классу. FPR, наоборот, ошибся и отнес профиль к неверному классу. Результаты тестирования моделей с различной функцией ядра представлены в табл. 1.

ЗАКЛЮЧЕНИЕ

Проведено тестирование метода опорных векторов для распознавания сложных и классических временных профилей микроволнового излучения на частоте 9.4 ГГц на основе данных каталога спектрополяриметра Нобеяма. Максимальная точность распознавания получена при радиальном ядре (RBF) и значении гамма равно единице. Алгоритм необходимо проверить на более значимой выборке данных и других частотных диапазонов.

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

Таблица 1.

Результат тестирования моделей SVM для распознавания временных профилей

	Linear (C=1, gamma=1)					RBF (C=10, gamma=1)					Polynomial (C=0.1, gamma=0.1)				
	TPR	FPR	Precision	Recall	F1	TPR	FPR	Precision	Recall	F1	TPR	FPR	Precision	Recall	F1
Классическая вспышка	62%	38%	0.73	0.92	0.81	85%	15%	0.79	0.85	0.81	92%	8%	0.67	0.50	0.77
Сложная вспышка	92%	8%	0.88	0.62	0.73	75%	25%	0.82	0.75	0.78	50%	50%	0.85	0.92	0.63

СПИСОК ЛИТЕРАТУРЫ

Davenport J.R.A., Hawley S.L., Hebb L. et al. Kepler Flares. II. The Temporal Morphology of White-light Flares on GJ 1243 // *Astrophysical J.* 2014. V. 797. iss. 2. 11 pp.
 Long Xu, Yi-Hua Yan, Xue-Xin Yu. et al. LSTM neural network for solar radio spectrum classification // *Res. Astronomy and Astrophysics.* 2019. V. 19, iss. 9. P. 135–147.
 Nishizuka N. Sugiura K., Kubo Y., Den M. et al. Solar Flare Prediction Model with Three Machine-learning Algorithms using Ultraviolet Brightening and Vector Magnetograms // *Astrophysical J.* 2017. V. 835, iss. 2. 14 pp.

Torii C., Tsukiji Y., Kobayashi S. et al. Full-automatic radiopolarimeters for solar patrol at microwave frequencies // *Proc. Research Institute of Atmospheric and Space Sciences, Nagoya University.* 1979. V. 26. P.129–132.
 Wang Y.H., Feng S.W., Du Q.F. et al. Solar Radio Burst Prediction Based on a Multimodal Model // *Solar Phys.* 2024. V. 299, iss. 4.
 URL: <https://solar.nro.nao.ac.jp/norp/html/event/> (дата обращения 23.05.2024).
 URL: <https://scikit-learn.org> (дата обращения 24.05.2024).