

Обработка материалов мультиспектральной съёмки на основе методов машинного обучения при управлении состоянием лесного массива

В. Ф. Мочалов

Санкт-Петербургский институт информатики
и автоматизации РАН
vicavia@yandex.ru

Р. С. Хабаров

Военно-космическая академия имени А. Ф. Можайского
xabarov1985@gmail.com

Аннотация. Рассматриваются вопросы применения методов машинного обучения при обработке материалов мультиспектральной космической съёмки. Выбор управляющего воздействия при управлении состоянием лесного массива осуществляется на основе регулярного анализа состояния лесной растительности. Идентификация вида лесной растительности и оценивание её состояния выполняются методами машинного обучения на базе спектрально-яркостных признаков. Предусматривается выбор обучающих и тестовых выборок. Представляются результаты сравнительного анализа методов машинного обучения при решении задач классификации и кластеризации лесных массивов для дальнейшего обоснованного выбора управляющего воздействия.

Ключевые слова: мультиспектральная космическая съёмка; спектрально-яркостные признаки, методы машинного обучения

1. ВВЕДЕНИЕ

Управление состоянием лесного массива осуществляется организациями, отвечающими за эксплуатацию леса. Ежегодно на территории России регистрируются несколько тысяч лесных пожаров. Актуальной задачей является разработка обоснованного управляющего воздействия на территорию. Рассматриваются следующие варианты реагирования на последствия лесного пожара: естественный, искусственный и комбинированный. При этом предполагается своевременная оценка таких основных показателей, как: площадь и вид сгоревшего растительного сообщества, категория лесного пожара, степень лесовозобновления после пожара. Решение задачи традиционными способами на основе данных наземных обследований связано с существенными материальными и временными затратами. В мировой практике при управлении состоянием лесного массива развиваются технологии автоматизированной обработки материалов мультиспектральной аэрокосмической съёмки [7–12]. При этом для повышения оперативности и достоверности активно совершенствуются технологии искусственного интеллекта. Вместе с тем при решении конкретной задачи с учетом информационных возможностей материалов

космической съёмки, спектрально-яркостных признаков для автоматизированной идентификации элементов ландшафта в условиях сезонной изменчивости требуется обосновать выбор оптимального метода автоматизированной обработки. В связи с этим представленное направление научных исследований имеет большую практическую значимость. Целью выполняемой работы является сравнительный анализ методов машинного обучения при обработке материалов мультиспектральной космической съёмки на примере одной из задач управления состоянием лесного массива.

Оптические свойства лесной растительности определяются способностью отражать, рассеивать, поглощать и излучать световую энергию. При распространении электромагнитных волн формируется сигнал, регистрируемый бортовой оптико-электронной мультиспектральной аппаратурой. После выполнения преобразований с учётом состояния атмосферы и условий распространения падающего и отражённого светового потока рассчитывается коэффициент спектральной яркости. Коэффициент спектральной яркости представляет собой физическую величину, характеризующую отношение яркости отражённого светового потока в определённом диапазоне длин волн в данном направлении к яркости абсолютно белой диффузно отражающей площадки при тех же условиях формирования сигнала. Расчёт коэффициента спектральной яркости обеспечивает возможность дальнейшего сравнительного анализа результатов спектрометрических измерений при использовании разновременных данных от различной бортовой и наземной спектрометрической аппаратуры [4].

Принимая во внимание случайный характер формирования значений коэффициентов спектральной яркости, целесообразно рассмотреть как традиционные методы машинного обучения, так и методы, основанные на математическом аппарате нечеткой кластеризации.

От выбора метода машинного обучения зависит качество управления состоянием лесного массива.

В качестве традиционных методов машинного обучения, обеспечивающих классификацию элементов ландшафта, рассматриваются, в частности, следующие методы: метод на основе персептрона, k -ближайших соседей, метод опорных векторов [1, 2] и др., а также их

комбинации. В качестве примеров кластеризации рассматриваются: метод k -средних и алгоритм нечётких C -средних или FCM-алгоритм [3].

Исходными данными для машинного обучения являются значения коэффициентов спектральной яркости в двенадцати спектральных каналах видимого и ближнего инфракрасного диапазонов спектра. Традиционные методы машинного обучения предполагают наличие дополнительных сведений о виде классифицируемого элемента ландшафта или так называемую метку класса. Результаты автоматизированной классификации и кластеризации проверяются на основе данных наземных обследований в контрольных точках.

II. ОСНОВНАЯ ЧАСТЬ

A. Исходные данные и методы обработки

В качестве примера в работе приводятся результаты обработки материалов мультиспектральной космической съёмки, полученной с помощью космического аппарата Sentinel-2 Европейского космического агентства [6]. Для формирования управляющего воздействия оцениваются последствия лесного пожара, произошедшего на территории Ленинградской области в 2019 году [14]. На рис. 1 приведена территория лесного пожара.

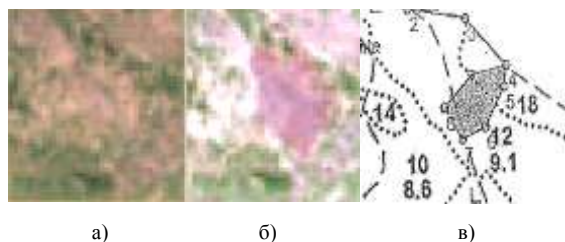


Рис. 1. Территория лесного пожара: а) – фрагмент материалов съёмки от 1 июня 2019 г. «до пожара»; б) – материалы съёмки от 2 июня 2020 г. «после пожара» в) – фрагмент схемы пожара (тёмный многоугольник) из акта о пожаре от 2 сентября 2019 г., составленного эксплуатирующей организацией

Предварительно на основе рассматриваемых методов машинного обучения в автоматизированном режиме проведено оценивание последствий лесного пожара. Были определены следующие показатели: площадь гари и вид сгоревшего растительного сообщества. Для оценивания точности определения показателей проведены наземные обследования.

В табл. I приведены исходные данные для машинного обучения. Каждая запись таблицы представляет собой результаты измерений, содержащие значения коэффициентов спектральной яркости в каждом из двенадцати спектральных каналов съемочной аппаратуры космического аппарата Sentinel-2 для соответствующих пикселей изображения, рис. 1, а также метку класса. Измерениям в пределах контура гари присвоена метка класса «1». Метки класса «2» и «3» присвоены участкам местности с типом лесной растительности 1 и 2 соответственно. Общее количество записей (пикселей изображения) в таблицах – 102 (по 34 записи для каждого класса).

ТАБЛИЦА I Исходные данные для машинного обучения

№ изм.	Номер канала						Метка класса
	1	2	3	...	11	12	
1	0.136	0.131	0.120		0.110	0.104	1
2	0.136	0.142	0.123		0.076	0.065	3
3	0.128	0.122	0.106		0.125	0.113	1
...							
100	0.140	0.110	0.122		0.081	0.066	3
101	0.140	0.132	0.129		0.092	0.075	3
102	0.125	0.118	0.099		0.087	0.075	2

В табл. II представлены результаты классификации с помощью традиционных методов машинного обучения [1, 2], а также их ансамблем (Ensemble).

ТАБЛИЦА II РЕЗУЛЬТАТЫ КЛАССИФИКАЦИИ

№ изм.	Метка класса	Методы классификации						
		PPN	SVM	Tree	LR	Forest	KNN	Ensemble
1	1	1	1	1	1	1	1	1
2	3	3	3	3	3	3	3	3
3	1	1	1	1	1	1	1	1
...								
100	3	3	3	3	3	3	3	3
101	3	3	3	3	3	3	3	3
102	2	2	2	2	2	2	2	2
accuracy		1,0	1,0	0,97	1,0	0,98	0,98	1,0

Методы классификации обозначены в табл. I следующим образом: метод на основе персептрона обозначен PPN, опорных векторов – SVM, «деревья решений» – Tree, логистическая регрессия – LR, «случайный лес» – Forest, k -ближайших соседей – KNN. Ансамбль реализован в виде мажоритарного голосования перечисленных выше методов. Метрика, обозначенная «ассигасу», является вероятностью правильной классификации, т. е. отношению числа правильно классифицированных тестовых измерений к общему числу измерений.

Значение метрики «ассигасу» для ряда методов оказалось равным 1,0, что обусловлено скорее недостаточно большим объемом исходных данных (измерений), чем качеством метода классификации. Для более точных оценок необходимо увеличить объем исходных данных. В табл. III представлены результаты кластеризации методами k -средних и FCM. Для расчета значений метрики «ассигасу» полученные метки кластеров были соотнесены с метками классов.

ТАБЛИЦА III РЕЗУЛЬТАТЫ КЛАСТЕРИЗАЦИИ

№ изм.	Метка класса	Метод кластеризации	Метод кластеризации			
			FCM			
			Меры принадлежности			Кл. max
		k средних	1	2	3	
1	1	1	2,20e-4	0,99	4,45e-5	2
2	3	2	2,57e-3	2,20e-4	0,99	3
3	1	1	3,05e-6	0,99	3,76e-8	2
...						
100	3	3	0,79	6,18e-4	0,21	1
101	3	2	5,46e-7	2,59e-8	0,99	3
102	2	3	0,99	2,35e-4	1,23e-4	1
accuracy		0,87	0,90			

Поскольку методы кластеризации k -средних и FCM являются методами обучения «без учителя»,

представленные в табл. III метки классов непосредственно не участвуют в процессе обучения. В обоих случаях количество кластеров указывается перед применением рассматриваемых методов машинного обучения.

Результатом работы метода k -средних являются метки кластеров с максимально различающимися значениями признаков. Результатом работы метода FCM являются меры принадлежности к каждому из кластеров. Для получения наиболее вероятной метки кластера следует выбрать номер с максимальной мерой принадлежности, что представлено в колонке «Кл. тах» табл. III.

Для получения истинных меток классов полученные метки кластеров можно соотнести с имеющимися в наличии маркированными данными. Предполагается, что число таких измерений невелико, иначе целесообразнее использовать методы обучения с учителем (классификации), дающими более точные результаты. Особенно актуально применение такого комбинированного подхода в случаях, когда присвоение меток классов связано с высокими материальными затратами.

На рис. 2 проиллюстрированы результаты работы методов классификации для новых данных, для которых отсутствуют метки классов. Затемненная область рисунка соответствует эталонной области сгоревшего леса, красным цветом выделены области, классифицированные моделями как «сгоревший лес», зеленым и синим – как лес типа 1 и 2 соответственно.

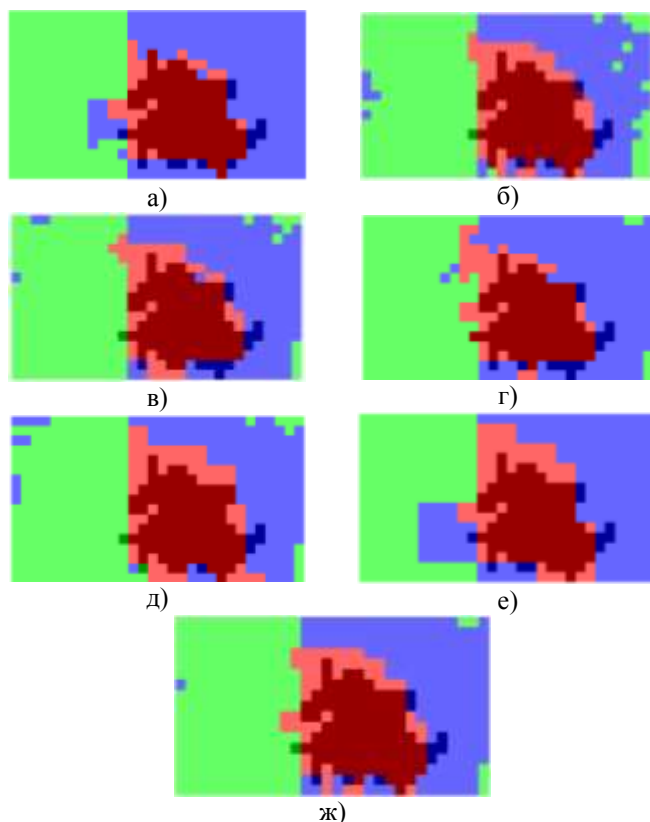


Рис. 2. Результаты работы методов классификации а) – случайный лес; б) – KNN; в) – логистическая регрессия LR; г) – SVM; д) – перцептрон; е) – деревья решений; ж) – ансамбль моделей

Как видно из рис. 1, наиболее точно повторяют контуры зоны гари методы SVM, ансамбль моделей и «случайный лес».

На рис. 3 представлены результаты работы рассмотренных методов кластеризации.

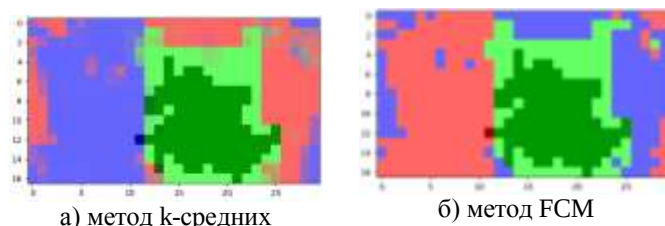


Рис. 3. Результаты работы методов кластеризации

Область, окрашенная тёмно-зеленым цветом, соответствует эталонной области гари. Для отображения результатов работы методов k -средних и FCM при кластеризации контура гари использовался светло-зеленый цвет.

III. ЗАКЛЮЧЕНИЕ

Анализ содержания данных, представленных в таблицах II и III, позволяет сделать выводы о точности оценивания показателей на основе рассматриваемых методов машинного обучения. Как и ожидалось, методы кластеризации идентифицируют элементы ландшафта с меньшей точностью, чем методы классификации. При этом необходимо учитывать, что методы кластеризации k -средних и FCM не предполагают построение обучающей выборки. Для их применения требуется указать лишь количество элементов ландшафта, подвергающихся кластеризации. Таким образом, требования к составу исходных данных, которые формируются, в основном в ходе наземных обследований, сокращаются, т. е. снижаются материальные затраты для их получения.

Следует также подчеркнуть, что управление состоянием лесного массива является многолетним процессом[5], в котором в качестве плановых операций осуществляется выбор управляющего воздействия и оценивание текущего состояния лесного массива.

СПИСОК ЛИТЕРАТУРЫ

- [1] Рашка С. Python и машинное обучение / пер. с англ. А.В. Логунова. М.: ДМК Пресс, 2017. 418 с.: ил.
- [2] Aurélien Géron. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow. Copyright © 2019 Kiwisoft S.A.S. 854 P.
- [3] Теребиж В.Ю. Введение в статистическую теорию обратных задач. М.: ФИЗМАТЛИТ, 2005. 376 с.
- [4] Шовенгердт Р.А. Дистанционное зондирование. Методы и модели обработки изображений. Техносфера, 2010. 560 с.
- [5] Потрясаев С.А., Соколов Б.В., Юсупов Р.М. Системно-управленческая интерпретация процессов создания и использования моделей и полимодельных комплексов // Информатизация и связь. 2019. № 3. С 14-19.
- [6] Sentinel-2 User Handbook ESA Standard Document Date 2015.
- [7] Airborne remote sensing. Theory and practice in assessment of terrestrial ecosystems. Semek F. et al Global Change Research Centre, 2014, 159 pages.

- [8] Olga Brovkina, Marko Stojanović, Slobodan Milanović, Iscander Latypov, Nenad Marković & Emil Cienciala Monitoring of post-fire forest scars in Serbia based on satellite Sentinel-2 data [GEOMATICS, NATURAL HAZARDS AND RISK] 2020, V 11, NO. 01, 2315–2339 <https://doi.org/10.1080/19475705.2020.1836037>
- [9] Anushree Badola, Santosh K. Panda, Dar A. Roberts, Christine F. Waigl, Uma S. Bhatt, Christopher W. Smith and Randi R. Jandt Hyperspectral Data Simulation (Sentinel-2 to AVIRIS-NG) for Improved Wildfire Fuel Mapping, Boreal Alaska [Remote Sens]. 2021, V 13, 1693. <https://doi.org/10.3390/rs13091693>
- [10] Kourosh Ahmadi, Bahareh Kalantar, Vahideh Saeidi, Elaheh K. G. Harandi, Saeid Janizadeh and Naonori Ueda Comparison of Machine Learning Methods for Mapping the Stand Characteristics of Temperate Forests Using Multi-Spectral Sentinel-2 Data [Remote Sens]. 2020, V 12(18), 3019; <https://doi.org/10.3390/rs12183019>
- [11] Mochalov V., Grigorieva O., Zhukov D., Markov A., Saidov A.: Remote Sensing Image Processing Based on Modified Fuzzy Algorithm // Proceedings of the 9th Computer Science On-line Conference, Vol.2 Springer Nature Switzerland AG 2020 R. Silhavy (Ed.): CSOC 2020, AISC 1225, pp. 563–572, 2020.
- [12] Мочалов В.Ф. Оценивание состояния растительности на основе аппарата нечёткой логики // Информатизация и связь. 2020. №5. С. 169-174. DOI: 10.34219/2078-8320-2020-11-5-169-174