

УДК 004.048

DOI: 10.18413/2518-1092-2020-5-2-0-3

**Кононов В.М.¹ | АЛГОРИТМ ПОДГОТОВКИ МУЛЬТИСПЕКТРАЛЬНЫХ
Асадуллаев Р.Г.² | СПУТНИКОВЫХ ДАННЫХ ДЛЯ ЗАДАЧИ КЛАССИФИКАЦИИ
Кузьменко Н.И.² | СЕЛЬСКОХОЗЯЙСТВЕННЫХ КУЛЬТУР**

¹) ООО «ЦентрПрограммСистем», ул. Восточная, д. 71, оф. 501, г. Белгород, 308019, Россия

²) Белгородский государственный национальный исследовательский университет,
ул. Победы, д. 85, г. Белгород, 308015, Россия

e-mail: kononov@lcp.ru, asadullaev@bsu.edu.ru, n.kuzmenko31@yandex.ru

Аннотация

в статье представлен разработанный алгоритм подготовки мультиспектральных спутниковых данных для задачи классификации сельскохозяйственных культур. Формализована процедура загрузки данных со спутника Copernicus Sentinel-2 с ресурса SentinelHub по заданным географическим координатам из geojson-файла с возможностью указания параметров, отражающих целевые даты получения данных. Описан алгоритм предобработки и аугментации данных к формату, необходимому для анализа математическими моделями машинного обучения.

Ключевые слова: классификация данных; машинное обучение; анализ многомерных данных; спутниковые снимки; сельскохозяйственные культуры; Copernicus Sentinel.

UDC 004.048

**Kononov V.M.¹ | ALGORITHM OF MULTI-SPECTRAL SATELLITE
Asadullaev R.G.² | DATA PREPARATION FOR AGRICULTURAL CROP
Kuzmenko N.I.² | CLASSIFICATION**

¹) «CentrProgrammSystem», LLC, 71, Vostochnaya St., Off. 501, Belgorod, 308019, Russia

²) Belgorod State National Research University, 85 Pobedy St., Belgorod, 308015, Russia

e-mail: kononov@lcp.ru, asadullaev@bsu.edu.ru, n.kuzmenko31@yandex.ru

Abstract

the article describes a developed algorithm of multispectral satellite data preprocessing for agricultural crop classification. The procedure for downloading data from the Copernicus Sentinel-2 satellite from the SentinelHub resource at the geographical coordinates from the shape file with ability to specify parameters that reflect the target data acquisition date is formalized. The algorithm of data preprocessing and augmentation to the format required for analysis by mathematical models of machine learning is described.

Keywords: data classification; machine learning; analysis of multidimensional data; satellite images; agricultural crops; Copernicus Sentinel.

ВВЕДЕНИЕ

На протяжении нескольких последних лет активно решается задача классификации сельскохозяйственных культур. Актуальность задачи обосновывается необходимостью ведения статистических данных о площади культур, выращиваемых в рассматриваемых регионах, расчета данных по продовольственному обеспечению государства, а также для осуществления мониторинга состояния посевов сельскохозяйственных культур.

Для этого используются различные методы и обрабатываются данные разного формата. В решении этой задачи можно использовать как снимки, полученные с беспилотных летательных аппаратов, так и мультиспектральные спутниковые снимки. Спутниковые данные применяются в

решении этой задачи в связи с простым и быстрым доступом к данным любой точки земной поверхности с небольшим временным интервалом между съемкой.

Сегодня для задачи классификации сельскохозяйственных культур чаще всего используются математические модели машинного обучения, такие как нейронные сети глубокого обучения (сверточные нейронные сети, сети долговременной памяти, рекуррентные нейронные сети и симбиоз данных архитектур), метод опорных векторов, случайный лес, метод k-средних, линейная регрессия и другие.

Для таких моделей в большинстве случаев необходимо приведение входных данных к одной размерности. Для повышения качества предсказаний модели используется предобработка изображений.

ОСНОВНАЯ ЧАСТЬ

Для задачи классификации сельскохозяйственных культур чаще всего используются следующие методы предобработки данных: приведение размеров снимков к единому размеру (256x256, 224x224, 128x128, 96x96), удаление фона, не относящегося к представленному классу, удаление пикселей со значением индекса NDVI ниже определенного порога, расчет дополнительных индексов вегетации и увлажненности с их конкатенацией к тренировочным примерам как дополнительные каналы, усреднение данных пикселей по значениям за год, вытягивание двумерного массива значений в одномерный вектор [3].

Для решения задач по классификации наземных объектов со спутниковых снимков важно высокое пространственное разрешение. С 2016 г. с периодичностью съемки в 10 дней стали доступны данные со спутника Copernicus Sentinel-2. Пространственное разрешение изображений Sentinel-2 в видимом и ближнем инфракрасном диапазонах спектра составляет 10 м, что лучше разрешения изображений другой серии спутников, Landsat (15 – 100 м). Второй спутник с аналогичными характеристиками Sentinel-2B был запущен 7 марта 2017 г., что позволило укоротить периодичность съемки до 5 дней [9].

Так, для решаемой задачи были выбраны данные со спутника Copernicus Sentinel-2. Было исследовано несколько способов загрузки спутниковых данных с данного спутника. Возможны следующие подходы:

- Copernicus Open Access Hub – сервис, поддерживающий скачивание тайлов (сплошных снимков размером 100x100 км). Данный продукт обладает ограниченным функционалом и не позволяет загружать отдельные области земной поверхности. Для сервиса также доступно API, в котором можно фильтровать снимки по общей облачности и дате съемки.

- SentinelHub – сервис, предоставляющий данные со спутников Sentinel, имеющий широкий инструментарий по интеграции с различными ГИС-программами и самостоятельно разрабатываемыми web-приложениями и программами. Сервис предоставляет несколько видов продуктов, таких как:

1. WMS API (Web Map Service) предоставляет простой HTTP-интерфейс для запроса географически привязанных изображений из одной или нескольких распределенных геопространственных баз данных. Запрос WMS определяет географические слои и область интереса, которая будет обработана. Ответом на запрос является одно или несколько географически зарегистрированных изображений карты (возвращаемых в формате JPEG, PNG и т. д.), которые могут отображаться в приложении браузера. Для настройки каналов и дополнительных данных, возвращаемых с сервера, используется конфигуратор.

2. Sentinel Hub Evalsript API – вместо использования конфигулятора возможно определять собственные пользовательские слои внутри Python и других языков программирования с помощью данного API. Для определения параметров запроса используется язык Javascript. После указания параметров они передаются на сервер в качестве HTTP-запроса. Данные с ресурса SentinelHub загружаются в виде многоканальных TIFF-изображений, где в каждом канале хранятся двумерные значения данных с 13 каналов с пространственным разрешением от 10 до 60 м на пиксель. [10]

В результате было выявлено, что для специфики задачи больше всего подходит загрузка данных с сервиса SentinelHub с помощью Evalsript API, так как этот сервис поддерживает запросы, в которых можно фильтровать облачность, выбирать загружаемые каналы и индексы вегетации, приводить скачиваемые данные к одному пространственному разрешению и, самое главное, присутствует возможность задавать географические координаты шейпов полей для выбора загружаемой области и автоматического заполнения нулевыми значениями до прямоугольной формы.

Для решения задачи классификации сельскохозяйственных культур были получены файлы формата geojson, содержащие уникальные идентификаторы полей, координаты точек, образующих форму полей и виды культур, засеянных за 2016 – 2019 год на этих полях.

При этом поля сильно различались в размерах и форме. Самые маленькие поля были площадью 4 га, а самые большие – 600 га. Было решено разработать алгоритм по приведению многоканальных снимков к одному размеру – 224x224 пикселя, при этом использовать различную логику для маленьких полей (до 150 пикселей), для средних полей (150 – 300 пикселей) и для больших полей (свыше 300 пикселей). Используемая схема изменения размеров представлена на рисунке 1.

Ширина/ высота	ширина меньше 150	ширина от 150 до 224	ширина от 225 до 300	ширина больше 300
высота меньше 150	изменение размера с сохранением пропорций			апсемплинг по узкой стороне и вырезание центра
высота от 150 до 224	доб. нулей по краям до размера 224		вырезание центра до размера 224x224	
высота от 225 до 300				
высота больше 300	апсемплинг по узкой стороне и вырезание центра		нарезка кусками 224x224 с переналожением	

Рис. 1. Схема изменения размеров изображений
Fig. 1. Images resizing scheme

Для скачивания и отбора данных для набора данных был разработан специальный алгоритм. Работа алгоритма начинается с загрузки и обработки файлов geojson. Для их обработки используется готовая библиотека, извлекающая нужные свойства по каждому полю и передающая их в массив. Далее в цикле происходит перебор дат с 1 апреля до 31 октября текущего года. Затем к дате добавляется 5 дней (временной интервал между снимками Sentinel-2) и формируется HTTP-запрос к серверу SentinelHub на получение спутниковых данных каналов B04-B08, B8A, B11 и B12 (видимые каналы RGB и каналы красного спектра) в заданном промежутке. После получения данных в TIFF-формате, они выгружаются в массив. Затем производится расчет среднего значения индекса вегетации NDVI по полю. Если это значение ниже порога в 0.6 или загруженное изображение слишком малой площади, данное изображение не сохраняется и происходит загрузка следующего. В ином случае массив данных сохраняется как файл в папку с данными классов культур. Алгоритм работы модуля загрузки и отбора спутниковых данных представлен на рисунке 2.

Следующим после загрузки данных шагом является их предобработка и аугментация. Для поставленной задачи решено было использовать следующие методы предобработки: нормализация данных к диапазону от 0 до 1 и приведение данных к нулевому среднему. Данные методы позволяют устранить большой разброс данных и привести данные к одному виду. Эти преобразования используются перед каждым запуском процесса обучения математической модели и получения предсказаний классов для всей выборки.

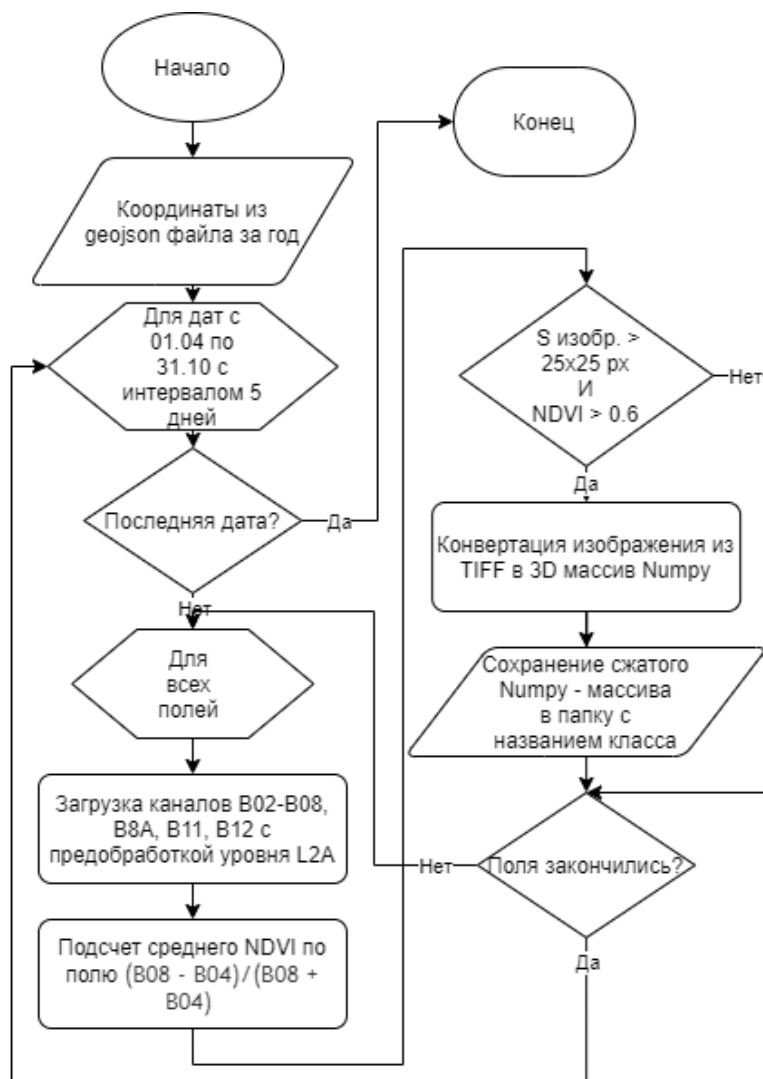


Рис. 2. Алгоритм работы модуля загрузки и отбора спутниковых данных
Fig. 2. The algorithm of the module for loading and selecting satellite data

Еще один метод, применяемый для предобработки – исключение из обучающей выборки изображений с высокой долей нулевых и облачных значений. Такой метод должен улучшить качество обучения.

Для обучения модели было решено использовать генератор – функцию, создающую аугментированные изображения, собирающую их в пакеты и подающую их на обучение в математическую модель. Таким образом, остается низкая вероятность появления в обучающих примерах исходных изображений, а это приводит к повышению качества обучения математической модели. Генератор позволяет решить такие проблемы как: неоднородность обучающей выборки, раннее переобучение модели, недостаточное число обучающих примеров и переполнение оперативной памяти при загрузке всего набора данных. Так как стандартный генератор используемой библиотеки машинного обучения Keras не поддерживает изображения с числом каналов больше 4, то необходимо было создать аналог такой функции. Алгоритм работы разработанного генератора представлен на рисунке 3. На рисунке 4 представлен пример работы генератора.

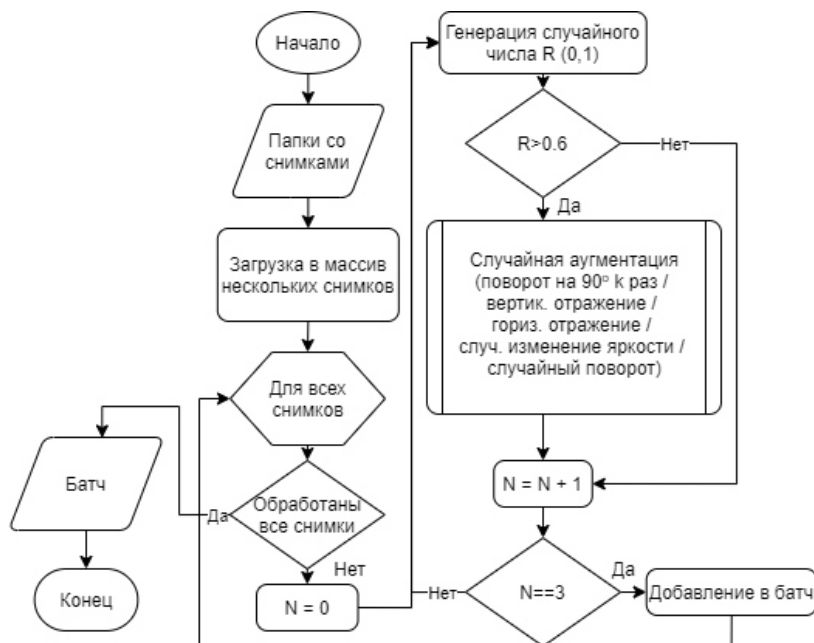


Рис. 3. Алгоритм работы генератора аугментированных изображений
Fig. 3. The algorithm of augmented images generator

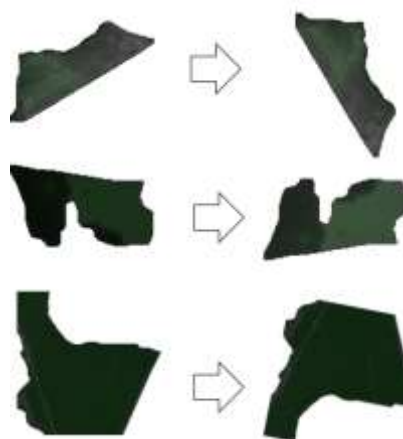


Рис. 4. Пример аугментаций изображений
Fig. 4. An example of the image augmentation

ЗАКЛЮЧЕНИЕ

Для задачи классификации сельскохозяйственных культур был разработан алгоритм подготовки мультиспектральных спутниковых данных. Алгоритм позволяет стандартизировать входные данные и повышает качество работы применяемых для классификации моделей машинного обучения за счет аугментации обучающей выборки и нормализации данных.

Также в ходе данной работы был разработан генератор аугментированных изображений, который позволяет решить проблему нехватки оперативной памяти при полной загрузке набора данных на обучение, а также позволяет перемешивать и аугментировать обучающие данные «на лету».

Разработанный алгоритм может быть полезным для тех, кто занимается обработкой спутниковых данных для их обработки с помощью математических моделей.

Список литературы

1. Комарова А. Ф., Журавлева И. В., Яблоков В. М. Открытые мультиспектральные данные и основные методы дистанционного зондирования в изучении растительного покрова // Принципы экологии. 2016. №1 (17). С. 40-74.
2. Герц Ж.В., Пулатов А.С., Миршадиев М.М. Пространственно-временная оценка покровных культур в Узбекистане с помощью дистанционного зондирования временных рядов // Актуальные вопросы науки. 2015. №22. С 66-75.
3. Чурсин И.Н., Филиппов Д.В., Горохова И.Н. Распознавание сельскохозяйственных культур по мультиспектральным космическим снимкам высокого разрешения // Вестник компьютерных и информационных технологий. 2018. №11 (173). С. 22-27.
4. Viskovic L., Kosovic I. N., Mastelic T. Crop Classification using Multi-spectral and Multitemporal Satellite Imagery with Machine Learning // 2019 International Conference on Software, Telecommunications and Computer Networks (SoftCOM), Split, Croatia, 2019. P. 1-5.
5. Rustowicz R. M. Crop Classification with Multi-Temporal Satellite Imagery // Stanford Project Posters and Reports, Fall 2017
6. Kamilaris A., Prenafeta-Boldú, F. X. Deep Learning in Agriculture: A Survey // Computers and Electronics in Agriculture. 2018. No. 147 (1). P. 70-90.
7. Shibendu R. Exploring machine learning classification algorithms for crop classification using Sentinel 2 data // ISPRS – International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. 2019. Vol. XLII-3/W6. P. 573-578.
8. Brandt J. Spatio-temporal crop classification of low-resolution satellite imagery with capsule layers and distributed attention. 2019. URL: <https://arxiv.org/pdf/1904.10130v1.pdf> (дата обращения: 15.04.2020)
9. The Copernicus Sentinel-2 mission URL: <https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-2> (дата обращения: 15.04.2020)
10. Sentinel Hub URL: <https://docs.sentinel-hub.com/> (дата обращения: 15.04.2020)

References

1. Komarova A., Zhuravleva I., Yablokov V. Open-source multispectral remote sensing data for the investigation of plant communities // Principy ekologii. 2016. No. 1 (17). P. 40–71.
2. Gercz Zh.V., Pulatov A.S., Mirshadiev M.M. Spatio-temporal assessment of cover crops in Uzbekistan using remote sensing time series // Aktualnye voprosy nauki. 2015. No. 22. P 66-75.
3. Chursin I. N., Philippov D. V., Gorokhova I. N. Practice in the recognition of crops on multispectral high-resolution satellite imagery // Vestnik kompiuternykh i informatsionnykh tekhnologii. 2018. No. 11 (173). P. 22-27.
4. Viskovic L., Kosovic I. N., Mastelic T. Crop Classification using Multi-spectral and Multitemporal Satellite Imagery with Machine Learning // 2019 International Conference on Software, Telecommunications and Computer Networks (SoftCOM), Split, Croatia, 2019. P. 1-5.
5. Rustowicz R. M. Crop Classification with Multi-Temporal Satellite Imagery // Stanford Project Posters and Reports, Fall 2017
6. Kamilaris A., Prenafeta-Boldú, F. X. Deep Learning in Agriculture: A Survey // Computers and Electronics in Agriculture. 2018. No. 147 (1). P. 70-90.
7. Shibendu R. Exploring machine learning classification algorithms for crop classification using Sentinel 2 data // ISPRS – International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. 2019. Vol. XLII-3/W6. P. 573-578.
8. Brandt J. Spatio-temporal crop classification of low-resolution satellite imagery with capsule layers and distributed attention. 2019. URL: <https://arxiv.org/pdf/1904.10130v1.pdf> (date of the application: 15.04.2020)
9. The Copernicus Sentinel-2 mission URL: <https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-2> (date of the application: 15.04.2020)
10. Sentinel Hub URL: <https://docs.sentinel-hub.com/> (date of the application: 15.04.2020)

Кононов Виктор Митрофанович, кандидат экономических наук, генеральный директор ООО «ЦентрПрограммСистем»

Асадуллаев Рустам Геннадьевич, кандидат технических наук, доцент, доцент кафедры прикладной информатики и информационных технологий

Кузьменко Николай Иванович, студент кафедры прикладной информатики и информационных технологий

Kononov Viktor Mitrofanovich, Candidate of Economical Sciences, General Director of «CentrProgrammSystem», LLC

Asadullaev Rustam Gennadievich, Candidate of Technical Sciences, Associate Professor of the Department of Applied Informatics and Information Technologies

Kuzmenko Nikolay Ivanovich, Bachelor Student, Department of Applied Informatics and Information Technologies