

## РАСПОЗНАВАНИЕ ПАХОТНЫХ ЗЕМЕЛЬ НА ОСНОВЕ ИЗМЕРЕНИЙ ВЕГЕТАЦИОННОГО ИНДЕКСА

С. А. Барталёв, А. Л. Загора

*Институт космических исследований Земли, г. Москва*

**Аннотация.** В работе представлен метод оценки используемости пахотных земель на основе измерений вегетационного индекса NDVI, вычисленных по изображениям спектрорадиометра MODIS со спутника Terra. Работа предполагает разработку программного модуля, выполняющего автоматическую классификацию сельскохозяйственных угодий на классы используемости. Разработан алгоритм и программный блок для классификации степени используемости сельскохозяйственных угодий. Проведено обучение классификатора с помощью алгоритмов машинного обучения. Для анализа использовались данные измерений вегетационного индекса NDVI за каждую неделю 2017 года, полученные в системе Vega аппаратом MODIS. Для тестирования алгоритмов использовалась выборка из 1026 векторов признаков (сельскохозяйственных угодий), равномерно распределенных по различным областям. Разбиение на обучающие и тестовые выборки производилось с сохранением баланса между областями. Результатом работы программы является метка класса для каждого входного вектора, характеризующая использование сельскохозяйственных угодий. Использование измерений вегетационного индекса позволило автоматизировать классификацию используемости полей. Для полей, вектора значений NDVI которых имеют пропуски измерений, применен алгоритм заполнения пропусков, что позволило использовать эти вектора для классификации. Были учтены особенности развития растительности в зависимости от географического положения. Использование алгоритмов машинного обучения дало следующие результаты: kNN – точность классификации 82%; SVC – точность классификации 78%; Random Forest – точность классификации 85%; GBT – точность классификации 86.307%. Наиболее высокие результаты получены с использованием алгоритма GBT - 86% правильно классифицированных полей. Программный блок реализован в виде модуля, готового для интеграции в систему Vega

**Ключевые слова:** ВЕГЕТАЦИОННЫЙ ИНДЕКС NDVI, КЛАССИФИКАЦИЯ СЕЛЬСКОХОЗЯЙСТВЕННЫХ УГОДИЙ.

**Анотація.** У роботі представлений метод оцінки використовуваного орних земель на основі вимірів вегетаційного індексу NDVI, обчислених із зображень спектрорадіометра MODIS з супутника Terra. Робота передбачає розробку програмного модуля, що виконує автоматичну класифікацію сільськогосподарських угідь на класи використовувані. Розроблено алгоритм і програмний блок для класифікації міри використовуваної сільськогосподарських угідь. Проведено навчання класифікатора за допомогою алгоритмів машинного навчання. Для аналізу використовувалися дані вимірів вегетаційного індексу NDVI за кожен тиждень 2017 року, отримані в системі Vega апаратом MODIS. Для тестування алгоритмів використовувалася вибірка з 1026 векторів ознак (сільськогосподарських угідь), рівномірно розподілених по різних областях. Розбиття на повчальні і тестові вибірки здійснювалося зі збереженням балансу між областями. Результатом роботи програми є мітка класу для кожного вхідного вектору, що характеризує використання сільськогосподарських угідь. Використання вимірів вегетаційного індексу дозволило автоматизувати класифікацію використовуваної полів. Для полів, вектору значень NDVI яких мають пропуски вимірів, застосований алгоритм заповнення пропусків, що дозволило використати ці вектору для класифікації. Були враховані особливості розвитку рослинності залежно від географічного положення. Використання алгоритмів машинного навчання дало наступні результати: kNN - точність класифікації 82%; SVC - точність класифікації 78%; Random Forest - точність класифікації 85%; GBT - точність класифікації 86.307%. Найбільш високі результати отримані з використанням алгоритму GBT - 86% правильно класифікованих полів. Програмний блок реалізована у вигляді модуля, готового для інтеграції в систему Vega

**Ключові слова:** ВЕГЕТАЦІЙНИЙ ІНДЕКС NDVI, КЛАСИФІКАЦІЯ СІЛЬСЬКОГОСПОДАРСЬКИХ УГІДЬ.

**Abstract.** The paper presents a method for assessing the utilization of arable land based on measurements of the NDVI vegetation index, calculated from images of a MODIS spectroradiometer from the Terra satellite. The work involves the development of a software module that performs the automatic classification of agricultural land into utilization classes. An algorithm and programmatic block are worked out for classification of degree of used of agricultural lands. Educating of classifier is conducted by means of algorithms of computer-aided instruction.

For an analysis data of measuring of vegetation index of NDVI were used for every week 2017 year, got in the system Vega the vehicle of MODIS. For testing of algorithms a selection was used from 1026 vectors of signs (agricultural lands) equipartition on different areas. Breaking up on teaching and test selections was produced with maintenance of balance between areas. A job of the program performance is a mark of class for every entrance vector, characterizing the use of agricultural lands. The use of measuring of vegetation index allowed to automatize classification of used of the fields. For the fields, vector of values of NDVI that is had admissions of measuring, the algorithm of filling of admissions is applied, that allowed to use these the vector for classification. The features of development of vegetation were taken into account depending on a geographical location. The use of algorithms of computer-aided instruction gave next results: kNN is exactness of classification of 82%; SVC is exactness of classification of 78%; Random Forest is exactness of classification of 85%; GBT is exactness of classification of 86.307%. The most high results are got with the use of algorithm of GBT - 86% of the correctly classified fields. Programmatic block realized as a module ready for integration in the system Vega

**Keywords:** VEGETATION INDEX NDVI, CLASSIFICATION OF AGRICULTURAL LAND.

## **Введение**

Одним из источников информации о состоянии сельскохозяйственных земель являются космические системы дистанционного зондирования Земли, позволяющие, в частности, проводить исследования, направленные на оценку состояния пахотных земель.

Выявлением задействованных в сельскохозяйственном обороте земель автоматизированными методами занимаются, в первую очередь, в Институте космических исследований РАН [1; 2]. Для классификации пахотных земель применяются многолетние временные ряды измерений спектральной яркости земной поверхности радиометром MODIS со спутника Terra. В качестве признаков распознавания используются различные вегетационные индексы, значение которых рассчитывается на основе коэффициентов спектральной яркости красного и ближнего инфракрасного диапазонов.

Самый используемый вегетационный индекс — индекс NDVI, однако формула вычислений вегетационных индексов имеет довольно много вариаций, предназначенных для уменьшения влияния различных помехообразующих факторов. В работе [2] приведены такие признаки распознавания пашни и даётся характеристика особенностей их использования.

В работе [3] описана методика распознавания объектов растительного покрова с использованием совместной обработки нескольких разносезонных космических изображений Landsat 8.

В представленной работе приведен алгоритм распознавания сельскохозяйственных угодий с использованием вегетационного индекса NDVI с использованием обучаемых алгоритмов классификации.

## **Постановка задачи**

Поставленная задача — создать программный продукт, цель которого построить алгоритм автоматической классификации сельскохозяйственных угодий по степени используемости по измерениям значения NDVI.

Измерения показателей NDVI проводились каждую неделю 2017 года, измерения представлены средним значением по всей площади поля.

Для того, чтобы формализовать понятие используемости поля будут использованы следующие классы: используемые; частично используемые; не используемые.

Класс частично используемых полей подразумевает, что поле принадлежит классу используемых и неиспользуемых одновременно.

В результате работы программного модуля каждому полю должен быть сопоставлен класс из вышеперечисленных.

Важно отметить, что программа рассчитана на интеграцию в систему Vega [6] и должна иметь строгий формат ввода и вывода.

Раскроем смысл понятий, используемых в дальнейшем. NDVI — нормализованный относительный индекс

растительности – показатель количества растительности (вегетационный индекс).

Расчет NDVI базируется на двух наиболее стабильных участках спектральной кривой отражения растений. В красной области спектра (0.6 – 0.7 мкм) лежит максимум поглощения солнечной радиации хлорофиллом высших растений, а в инфракрасной области (0.7 – 1.0 мкм) находится область максимального отражения клеточных структур листа. То есть высокая фотосинтетическая активность ведет к меньшему отражению в красной области спектра и большему в инфракрасной. Отношение этих показателей друг к другу позволяет четко отделять и анализировать растительные от прочих природных объектов. Использование же нормализованной разности между минимумом и максимумом отражений увеличивает точность измерения, позволяет уменьшить влияние таких явлений как различия в освещенности снимка, облачности и пр.

Для вычислений вегетационного индекса NDVI используются каналы красной области спектра и ближней инфракрасной области спектра. Индекс NDVI вычисляется по следующей формуле:

$$NDVI = \frac{NIR - RED}{NIR + RED},$$

где NIR — отражательная способность в ближней инфракрасной области спектра;

RED – отражательная способность в видимой красной области спектра. На выходе получается изображение со значениями пикселей, изменяющихся от -1 до 1.

Нормализованный вегетационный индекс (NDVI) – это стандартизированный индекс, показывающий наличие и состояние растительности (относительную биомассу). Этот индекс использует контраст характеристик двух каналов из набора мультиспектральных растровых данных – поглощения пигментом хлорофилла в красном канале и высокой

отражательной способности растительного сырья в инфракрасном канале (NIR).

Здоровая растительность обычно показывает лучшее отражение в ближнем инфракрасном диапазоне, чем в красной области видимого спектра. Инфракрасный диапазон излучения поглощается облаками, водой и снегом, и отражается камнями и голой почвой почти так же, как диапазон красной области видимого спектра. NDVI используется по всему миру для мониторинга засухи, мониторинга и прогнозирования сельскохозяйственного производства и картографирования наступления пустыни. NDVI предпочтительнее для глобального мониторинга растительности, поскольку помогает компенсировать изменение условий освещения, уклон поверхности, экспозицию и другие внешние факторы.

Очень маленькие значения (< 0.1) измерений NDVI соответствуют пустым областям скал, песка или снега. Умеренные значения (от 0.2 до 0.3) представляют кустарники и луга, в то время как большие значения (от 0.6 до 1.0) указывают на умеренные и тропические леса.

Отметим, что в работе [2] приведены также и другие признаки распознавания пашни и особенности их использования.

### Решение задачи

Для достижения поставленной цели в работе решаются следующие задачи:

- получение данных;
- обработка данных, получение признаков;
- построение классификатора на основе полученных данных.

Также нужно сформулировать технические требования для реализуемой программы.

Требования к функциональности программы таковы:

- реализовать алгоритм построения классификатора векторов признаков сельскохозяйственных угодий;
- каждому экземпляру данных классификатор должен сопоставить метку: 1 – используется; 2 – частично используется; 3 – не используется;

- программа должна выполнять роль модуля, готового для интеграции в систему Vega [6];

- модуль должен корректно обрабатывать ошибки, в случае ошибки возвращать информацию, понятную пользователю. Признаки, характерные различным типам сельскохозяйственных угодий.

Яровые культуры. Весной, в начале мая, поздние яровые и пропашные культуры ещё не посеяны, а ранние яровые культуры находятся на стадии всходов и проективное покрытие почвы на полях под этими культурами очень низкое.

На момент посева на полях всех яровых культур наблюдаются значения индекса NDVI, характерные для открытой почвы. После уборки урожая (август, сентябрь) поля с открытой почвой принимают оттенки от розового до коричневого и хорошо отличаются от естественной, часто еще вегетирующей, растительности.

Озимые культуры. Особенность – раннее начало вегетации. Хорошо идентифицируются на весенних снимках, особенно до начала периода вегетации, и на снимках, сделанных поздней осенью, когда вегетационный период естественной травянистой растительности уже закончен. Для того, чтобы отличить озимые культуры от яровых следует пользоваться серией снимков, поскольку, например, поля под паром в весенний период могут быть определены как яровые, а некоторые многолетние травы – как озимые. Осенний снимок (август-октябрь) позволит увидеть, что на полях с озимыми культурами прошел сбор урожая и значения индекса NDVI низкие, а у большинства культур наблюдается вегетационная фаза созревания.

Поля под паром. Сезонная динамика индекса NDVI на полях под паром слабо изменяется (небольшие пики обусловлены, как правило, ростом сорняков) и имеет очень низкий объем фитомассы (значения NDVI редко поднимаются выше 0.5); Поля под чистым паром имеют яркость и цвет, присущий открытой почве (коричнево-

розовые оттенки при использовании синтеза каналов RED – NIR – GREEN ИЛИ SWIR1 – NIR – RED).

Неиспользуемые и частично используемые земли. Неоднородная мозаичная структура, которая образуется растительностью или почвенным покровом неодинаковой влажности и характерна для беспорядочно расположенных участков различного тона, размеров и форм; Характеризуется такими же цветовыми и текстурными признаками, что и естественные луга, однако схожа по форме с используемыми пахотными полями. Особенно это справедливо для полей, заброшенных недавно.

Для идентификации залежей необходим анализ серии снимков.

Описание практической части. Для реализации поставленной задачи был выбран язык Python [4;5] и среда разработки Pycharm. Программа реализована в виде модуля, готового для интеграции в систему Vega [6;7]. Для работы с данными выбрана библиотека pandas. Выбор библиотеки обусловлен тем, что в ней реализовано большинство необходимых известных алгоритмов обработки данных, а также стоит отметить ее высокую производительность. Для реализации алгоритмов машинного обучения использована библиотека scikit-learn.

Подготовка данных. На вход программному модулю подаются вектора измерений вегетационного индекса NDVI за каждую неделю 2017 года, полученные в системе Vega аппаратом MODIS. Основная проблема на этом этапе заключается в том, что некоторые экземпляры данных имеют пропуски в измерениях. Для того, чтобы выровнять данные, были разработаны алгоритмы, позволяющие заполнить недостающие значения.

Был использован следующий способ – пропуски заполнялись средним значением NDVI для полей по текущей неделе (среднее высчитывалось по тем полям, у которых есть измерение на этой неделе). При отсутствии значений для всех полей, эти пропуски заполнялись соседними

значениями. В результате работы алгоритма все вектора значений NDVI имеют равную длину и пригодны для использования в разных областях. При этом, графики изменения вегетационного индекса NDVI для сельскохозяйственных полей в северной и в южной частях, даже при условии того, что на них выращивается одна и та же культура, имеют существенные отличия друг от друга.

Что бы учесть фактор географического положения, было принято решение к вектору признаков добавить признак области, в котором находится объект (угодье).

Используемые библиотеки, в частности, библиотека scikitlearn напрямую не обрабатывает категориальные признаки. Поэтому прежде чем подавать данные на вход алгоритмов машинного обучения необходимо преобразовать категориальные признаки в количественные. Был использован метод векторизации, суть которого заключается в следующем:

Признак  $j$ , принимающий  $s$  значений, заменим на  $s$  признаков, принимающих значения 0 или 1, в зависимости от того, чему равно значение исходного признака  $j$ .

Таким образом, к вектору признаков добавляется 25 столбцов, где на позиции нужной области стоит единица. Результат показал, что добавление этих признаков оправдано, точность классификации существенно увеличилась.

Обучение классификатора проводилось с помощью 4-х алгоритмов:

- kNN – метод ближайших соседей;
- SVC – машина опорных векторов;
- Random Forest – случайный лес;
- GBT – градиентный бустинг

деревьев решений.

Для тестирования алгоритмов использовалась выборка из 1026 векторов признаков (сельскохозяйственных угодий), равномерно распределенных по различным областям. Для обучения и тестирования классификатора исходная выборка была разбита на обучающую и тестовую в соотношении 77% и 33% соответственно. Так же стоит отметить, что разбиение на обучающие и тестовые выборки

производилось с сохранением баланса между областями. После разбиения эти вектора были использованы как входные данные для алгоритмов. Результатом работы программы является метка класса для каждого входного вектора: 1 – используется, 2 – частично используется, 3 – не используется.

В результате настройки и обучения классификатора были получены следующие результаты для указанных входных данных:

1. kNN – точность классификации 82%;
2. SVC – точность классификации 78%, все поля были классифицированы как используемые, что говорит о плохой обучаемости этого алгоритма на входных данных;
3. Random Forest – точность классификации 85%, достаточно точно классифицированы частично используемые поля (91% для этого класса);
4. GBT – лидирующий алгоритм, точность классификации 86.307%.

Заключение. В работе рассмотрены факторы, определяющие степень используемости сельскохозяйственного угодья.

Разработан алгоритм для классификации степени используемости сельскохозяйственных угодий, а также создана программа, которая реализует разработанный алгоритм.

Программа написана на языке Python с использованием среды разработки Pycharm, библиотек pandas и scikit-learn.

### Выводы

1. Использование измерений вегетационного индекса позволило автоматизировать классификацию используемости полей;
2. Для полей, вектора значений NDVI которых имеют пропуски измерений, применен алгоритм заполнения пропусков, что позволило использовать эти вектора для классификации;

3. Были учтены особенности развития растительности в зависимости от географического положения;
4. Использование алгоритмов машинного обучения дало следующие результаты: kNN – точность классификации 82%; SVC – точность классификации 78%; Random Forest – точность классификации 85%; GBT – точность классификации 86.307%;
5. Наиболее высокие результаты получены с использованием алгоритма GBT-86% правильно классифицированных полей.

### Библиографические ссылки

1. Барталев С.А., Егоров В.А., Жарко В.О., Лупян Е.А., Плотников Д.Е., Хвостиков С.А., Шабанов Н.В. Спутниковое картографирование растительного покрова России: М. ИКИ РАН, 2016. 208 с.
2. Барталев С.А., Егоров В.А., Лупян Е.А., Плотников Д.Е., Уваров И.А. Распознавание пахотных земель на основе многолетних спутниковых данных спектрорадиометра MODIS и локально-адаптивной классификации. Компьютерная оптика, 2011. Т. 35. № 1. С. 103–116.
3. Стыценко Е.А. Возможности распознавания сельскохозяйственных угодий с использованием методики совместной автоматизированной обработки разносезонных многозональных космических изображений. Современные проблемы дистанционного зондирования Земли из космоса, 2017. Т. 14. № 5. С. 172–183
4. Лутц, М. Программирование на Python, II том: М. Лутц. - СПб: Символ-плюс, 2015.

5. Max Kuhn, Kjell Johnson. Applied Predictive Modeling: Springer, 2019.
6. ВЕГА-Science [Электронный ресурс]: URL: <http://sci-vega.ru>.
7. Загора А.Л., Барталев С.А. Распознавание пахотных земель на основе измерения индекса NDVI //Материалы XXI Международной молодежной научно-практической конференции «Человек и Космос» (10 - 12 апр. 2019г.) - Днепр: Национальный центр аэрокосмического образования молодежи им. Макарова, 2019. - DVD, ISSN 2221-4550.

*Надійшла до редколегії 30.05.2019р.*

### Сведения об авторах



Барталёв Сергей Александрович, Россия. Институт космических исследований Земли Заведующий лабораторией спутникового мониторинга наземных экосистем д.т.н., профессор. Сфера интересов – дистанционное зондирование Земли



Загора Александр Леонидович, Россия. Институт космических исследований Земли г. Москва. Аспирант. Сфера интересов – дистанционное зондирование Земли