

# КЛАССИФИКАЦИЯ СЕЛЬСКОХОЗЯЙСТВЕННЫХ КУЛЬТУР НА ОСНОВЕ АНАЛИЗА ВРЕМЕННЫХ РЯДОВ ВЕГЕТАЦИОННОГО ИНДЕКСА С ПОНИЖЕНИЕМ ИХ РАЗМЕРНОСТИ

© 2023 г. М. А. Павлова<sup>1,\*</sup>, Д. С. Сидорчук<sup>1</sup>, Д. А. Бочаров<sup>1</sup>

<sup>1</sup> Институт проблем передачи информации им. А.А. Харкевича РАН  
127051 Москва, Большой Каретный пер., 19, Россия

\*E-mail: pavlovamar96@gmail.com

Поступила в редакцию 10.03.2023 г.

После доработки 28.03.2023 г.

Принята к публикации 29.03.2023 г.

В работе рассматривается задача классификации сельскохозяйственных культур. Как известно, для решения этой задачи значительно эффективнее использовать не только мгновенные данные дистанционного зондирования или вычисленные по ним вегетационные индексы, но и их исторический набор для разных моментов времени. Временные ряды, образованные значениями индексов для фиксированной пространственной точки для разных моментов времени, характеризуются высоким уровнем пропусков значений, вызванных в первую очередь наличием облачности в некоторые даты. Проведено исследование известных методов аппроксимации временных рядов. Также исследуется вопрос о том, может ли снижение размерности аппроксимированных временных рядов повысить качество рассматриваемой классификации. В экспериментальной части работы использовались временные ряды индекса NDVI, вычисленного по мультиспектральным спутниковым данным Sentinel-2. Исследовалась классификация кукурузы, подсолнечника, пшеницы и сои. В работе показано, что снижение размерности методом UMAP позволяет в среднем в 1.5 раза повысить значение F1-меры в сравнении с использованием данных исходной размерности. Предложен новый метод классификации культур, основанный на аппроксимации временных рядов вегетационных индексов NDVI кубическим сплайном, извлечение малоразмерных признаков алгоритмом UMAP и их классификации методом  $k$  ближайших соседей.

**Ключевые слова:** ДЗЗ, классификация сельскохозяйственных культур, временные ряды, NDVI, аппроксимация временного ряда, извлечение признаков, снижение размерности, UMAP

**DOI:** 10.31857/S023500922302004X, **EDN:** QSZZFT

## ВВЕДЕНИЕ

В последние годы благодаря развитию технологий хранения и обработки больших массивов данных в сельском хозяйстве для мониторинга состояния земель все активнее применяются спутниковые данные дистанционного зондирования Земли (ДЗЗ) (Якушев и др., 2019). Современные программы космического зондирования позволяют получать актуальную информацию о поверхности Земли с большой частотой и высоким пространственным разрешением, что обогащает данные, получаемые наземными методами мониторинга. Данные ДЗЗ активно применяются в рамках точного или цифрового земледелия (Блохина, 2018; Бахтадзе и др., 2020) — концепции эффективного управления сельскохозяйственным производством с применением современных информационных технологий.

На данный момент активно развиваются методы, позволяющие по спутниковым данным составлять и обновлять карты границ сельскохозяйственных угодий, оценивать и прогнозировать урожайность, картографировать местность (Sishodia et al., 2020). Особенно актуальным является развитие методов и технологий автоматического определения типов сельскохозяйственных культур, поскольку информация об их пространственных распределениях критически важна в эффективном сельскохозяйственном управлении и производстве (Orynbaikyzy et al., 2019). В дополнение автоматическая классификация культур может способствовать совершенствованию методов эквализации условий съемки, основанных на использовании специальных зон изображения, для которых известны априорные предположения о спектральных функциях отражения поверхности (Павлова и др., 2022).

Известно, что отражательная способность растительности существенно различается для разных длин волн спектра падающего излучения (Черепанов, Дружинина, 2009). С использованием наборов различных каналов мультиспектральных спутниковых снимков получают так называемые спектральные индексы земной поверхности. Те индексы, что применяются для анализа состояния растительности, известны как вегетационные индексы.

Особенности развития различных растений оказывают влияние на динамику их спектральных показателей. В работах (Барталев и др., 2006; Пугачева, Шевырнов, 2008; Sun et al., 2019; Плотников и др., 2011) показана разделимость сельскохозяйственных культур на основе их спектрально-временных профилей, в том числе на ранних этапах вегетации. Культуры, имеющие сходные спектральные характеристики на одном временном отрезке и потому неотличимые, могут иметь значительные спектральные различия на другом временном отрезке (Mugni, Biswas, 2015). Таким образом, задачу классификации сельскохозяйственных культур значительно эффективнее решать, используя временные ряды или серии значений, а не мгновенные значения.

В работе (Sun et al., 2019; Плотников и др., 2011) показано, что увеличение числа наблюдений в период активной вегетации улучшает качество классификации. Однако объем доступных данных часто бывает ограничен из-за облачности. Если данные, попавшие в область тени от облаков, могут быть с некоторой точностью восстановлены (Бочаров и др., 2021), то данные в точках областей плотной облачности использоваться не могут. Во многих работах (Zhang et al., 2011; Rußwurm, Körner, 2020; Воробьева, Чернов, 2017; Hird, McDermid, 2009) эту проблему решают при помощи аппроксимации значений временного ряда. В работе (Воробьева, Чернов, 2017) сплайновая модель показала наилучшее качество в сравнении с рядом альтернативных моделей. В настоящей работе мы добавляем к сравнению регрессию на основе гауссовских процессов, не рассматривавшуюся в работе Воробьевой. Данный вид аппроксимации исследовался в работе (Belda et al., 2020) как перспективный метод обработки пропусков в данных.

Известно, что извлечение признаков, содержащих наибольшее количество полезной информации, из входных данных и классификация на их основе значительно повышают качество относительно классификации по необработанным данным (Velliangiri et al., 2019). Используя такой подход, можно не только снизить вероятность переобучения, но и уменьшить время работы классификатора. Однако выбор оптимальных призна-

ков для обучения классификатора является не тривиальной задачей (Abe, Jordaаn, 2013).

В работе (Yang et al., 2020) предлагается формировать признаковое пространство из различных спектральных и текстурных характеристик, рассчитанных для трех экспертно выбранных дат, а затем выбирать оптимальные признаки при помощи одного из рассматриваемых в работе алгоритмов. В материале (Пугачева, Шевырнов, 2008) предлагаются другие экспертные признаки, а именно углы наклона кривой NDVI в период роста и созревания, а также максимальное значение индекса за весь вегетационный период, таким образом переходя от временного ряда NDVI к трем значениям.

В настоящей работе мы также развиваем классификацию за счет извлечения признаков. В качестве исходного признакового пространства, также, как и в работе (Пугачева, Шевырнов, 2008), используется временной ряд NDVI. В отличие от двух названных выше работ, применяющих извлечение признаков, мы не предлагаем новых экспертных признаков. Вместо этого для извлечения признаков временного ряда мы предлагаем использовать алгоритм снижения размерности UMAP (McInnes et al., 2018). Такой алгоритм уже применялся в области классификации сельскохозяйственных культур в работе (Rußwurm, Körner, 2020). Однако это применение сводилось к кластеризации для разведочного анализа данных, а именно, чтобы установить саму возможность классификации. Снижение размерности в качестве предварительного этапа перед классификацией предлагалось в работе (Gilbertson, Van Niekerk, 2017), где авторы рассматривают в качестве исходного признакового пространства не временной ряд, а различные пространственные и спектральные характеристики.

Собственно классификацию культур в данной области выполняют с применением различных моделей. В работе (Chakhar et al., 2021) исследовались методы классификации на основе комбинаций радиолокационных данных от спутника Sentinel1 и вегетационных индексов, вычисленным по оптическим данным Sentinel2. Наилучшие показатели качества были получены с использованием машины опорных векторов (cubic SVM). Ли и соавт. в своей работе (Li et al., 2020) предложили новую генеративную нейросетевую модель, позволяющую классифицировать культуры на три класса (соя, кукуруза и другие) по мультивременным спутниковым данным, состоящим из трех дат. В работе (Reedha et al., 2022) предложена классифицирующая нейросетевая модель на основе визуального трансформера (visual transformer). В работе (Sun et al., 2019) для классификации применялись методы SVM и RF (random forest). Авторами в работе (Фирсов и др., 2021) для клас-

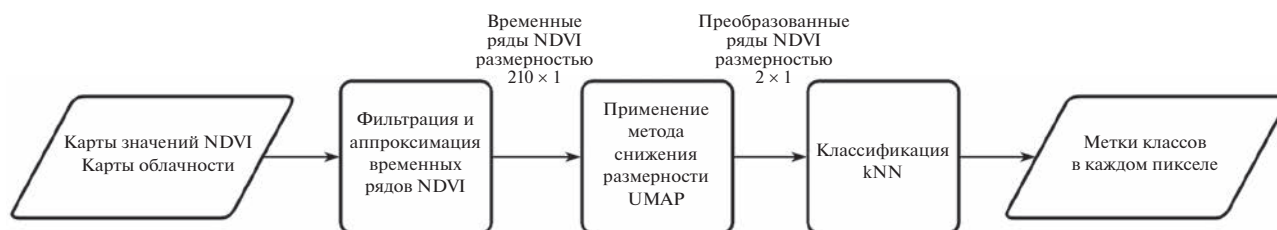


Рис. 1. Блок-схема предлагаемого метода классификации.

сификации гиперспектральных данных предложена спектрально-пространственная сверточная нейронная сеть. В работе (Gilbertson, Van Niekerk, 2017) в качестве классификаторов исследовались методы SVM, деревья решений (DT), метод k-ближайших соседей (k-NN), а также метод случайного леса (RF). Для последних трех методов было заявлено улучшение качества при добавлении предварительного этапа извлечения признаков. В нашей работе в качестве классификатора используется алгоритм kNN, превзошедший алгоритмы SVM и DT в другой задаче классификации ДЗЗ снимков, согласно результатам работы (Bouteldja, Kourgli, 2020).

## ЭКСПЕРИМЕНТАЛЬНЫЕ ДАННЫЕ

Нами исследовались временные ряды вегетационного индекса NDVI, состоящие из 40 наблюдений. Каждое наблюдение представляло собой пару из изображения индекса размером  $5730 \times 5730$  пикселей с пространственным разрешением 10 м на пиксель, рассчитанного по данным спутниковых мультиспектральных снимков Sentinel-2, и карты облачности, используемой для исключения областей неопределенности данных.

Обучение классификатора и оценку качества алгоритма проводили на основе данных о землепользовании в Краснодарском крае Российской Федерации, представленных в виде координат границ земельного участка и соответствующей метке класса сельскохозяйственной культуры. Данные содержат следующие метки классов: “Кукуруза”, “Подсолнечник”, “Пшеница”, “Соя”, “Другое”. В класс “Другое” на выбранных изображениях попали невозделываемые поля, лесопарковые зоны, участки дорог и городской застройки. Тестовая выборка включала по одному полю каждого класса, остальные поля использовались для обучения.

## ПРЕДЛАГАЕМЫЙ МЕТОД

Предлагаемый метод классификации состоял из следующих основных шагов (рис. 1): аппроксимации временных рядов вегетационного индекса; снижении размерности пространства дан-

ных; классификации временных рядов пониженной размерности.

На первом шаге заполняли пропуски временных рядов индекса NDVI, вызванные облачностью, удаляли шумы и генерировали равномерную сетку значений, что достигали путем аппроксимации временных рядов. Для извлечения признаков из аппроксимированных временных рядов предлагали использовать алгоритм UMAP с размерностью конечного пространства признаков, равной двум.

Далее в экспериментальной части работы исследовали разные модели аппроксимации временных рядов и проводили сравнение качества предлагаемого метода с его модификацией, использующей для извлечения признаков метод главных компонент.

## ЭКСПЕРИМЕНТЫ

В этом разделе описывается методика двух экспериментов. Первый эксперимент нацелен на сравнение различных моделей аппроксимации временных рядов вегетационного индекса. Второй эксперимент посвящен оценке эффекта от использования снижения размерности в качестве предварительного этапа классификации временных рядов NDVI.

### *Оценка эффективности аппроксимации временных рядов вегетационного индекса*

Цель эксперимента — выбор оптимальной математической модели, наиболее точно описывающей поведение вегетационного индекса в условиях пропусков в данных. Рассматривались следующие модели: полиномы второй и третьей степени (Groten, 1993; Воробьева, Чернов, 2017), модель сплайна третьего порядка (Воробьева, Чернов, 2017) и регрессия на основе гауссовских процессов (Belda et al., 2020).

Для выбора оптимальной математической модели исследуемых в данной работе значений индекса были приняты следующие критерии качества аппроксимации:

### Среднеквадратическая ошибка

Среднеквадратическая ошибка говорит об усредненном качестве аппроксимации и рассчитывается по формуле:

$$MSE = \frac{1}{N} \sum_{i=1}^N (Y_i - \hat{Y}_i)^2,$$

где  $Y_i, \hat{Y}_i$  — истинное и оценочное значения индекса NDVI в  $i$ -й точке,  $N$  — количество точек.

### Максимальная ошибка

Максимальная ошибка дает представление о максимальных отклонениях значений NDVI от аппроксимирующей функции, что позволяет оценить применимость данной функции к рассматриваемой предметной области. В данной работе значение максимальной ошибки оценивалось как 99-й перцентиль всех отклонений оценок временных рядов NDVI от их истинных значений.

### Среднеквадратическая ошибка воспроизводимости

Среднеквадратическая ошибка воспроизводимости позволяет оценить устойчивость аппроксимирующей функции к пропускам в данных, вызванным облачностью и тенями от облаков. За ошибку воспроизводимости была принята величина, рассчитываемая как попарная разница значений в одной и той же точке, полученная при выполнении аппроксимации временного ряда несколько раз с оптимизацией на случайном наборе точек этого ряда.

Среднеквадратическая ошибка воспроизводимости рассчитывается по следующей формуле:

$$MSE_{\text{воспр.}} = \frac{1}{N(G-1)} \sum_{n=1}^N \sum_{i=1}^G \sum_{j=i+1}^G (\hat{Y}_{n,i} - \hat{Y}_{n,j})^2,$$

где  $G$  — количество повторов аппроксимаций в одной точке,  $N$  — количество точек, для которых производился расчет,  $\hat{Y}_{n,i}, \hat{Y}_{n,j}$  — оценки значений  $i$ -й и  $j$ -й аппроксимаций в точке  $n$ .

Производительность выбранного алгоритма аппроксимации также является важной характеристикой, так как планируется применять разработанный метод в экспертных системах. Среднее время обработки ряда рассчитывалось следующим образом:

$$\overline{\text{time}} = \frac{1}{N} \sum_{i=1}^N \text{time}_i,$$

где  $\text{time}_i$  — время выполнения  $i$ -й аппроксимации,  $N$  — количество выполненных аппроксимаций.

Эксперимент проводился на наборе временных рядов, представленных в виде двумерной

матрицы, где каждая строка представляет собой измерения вегетационного индекса NDVI за 2018 календарный год.

Для симуляции потери данных ввиду облачности из рядов случайным образом исключается некоторая доля точек. Исследовалось поведение аппроксимирующих функций при долях пропущенных значений, равных 0.2, 0.33 и 0.5. Затем значения индекса аппроксимируются исследуемыми моделями, после чего проводится сравнение описанных характеристик полученных моделей и выбирается наиболее оптимальная модель с точки зрения точности аппроксимации и производительности.

Оптимизация предложенных моделей производилась методом наименьших квадратов. Все характеристики оценивались для каждой рассматриваемой доли пропущенных значений. Для оценки среднеквадратической ошибки воспроизводимости каждый ряд генерировался 10 раз на основе случайного набора точек из исходной последовательности в соответствии с текущей долей пропущенных значений.

### Сравнение методов снижения размерности

Цель данного эксперимента — сравнение влияния алгоритмов понижения размерности на качество классификации, среди которых рассмотрены PCA и UMAP. В качестве реализации алгоритма UMAP использовалась библиотека umap-learn на языке Python3 со следующими параметрами:  $n\_neighbors=200$ ,  $min\_dist=0.1$ ,  $n\_components=2$ ,  $metric = 'euclidean'$ .

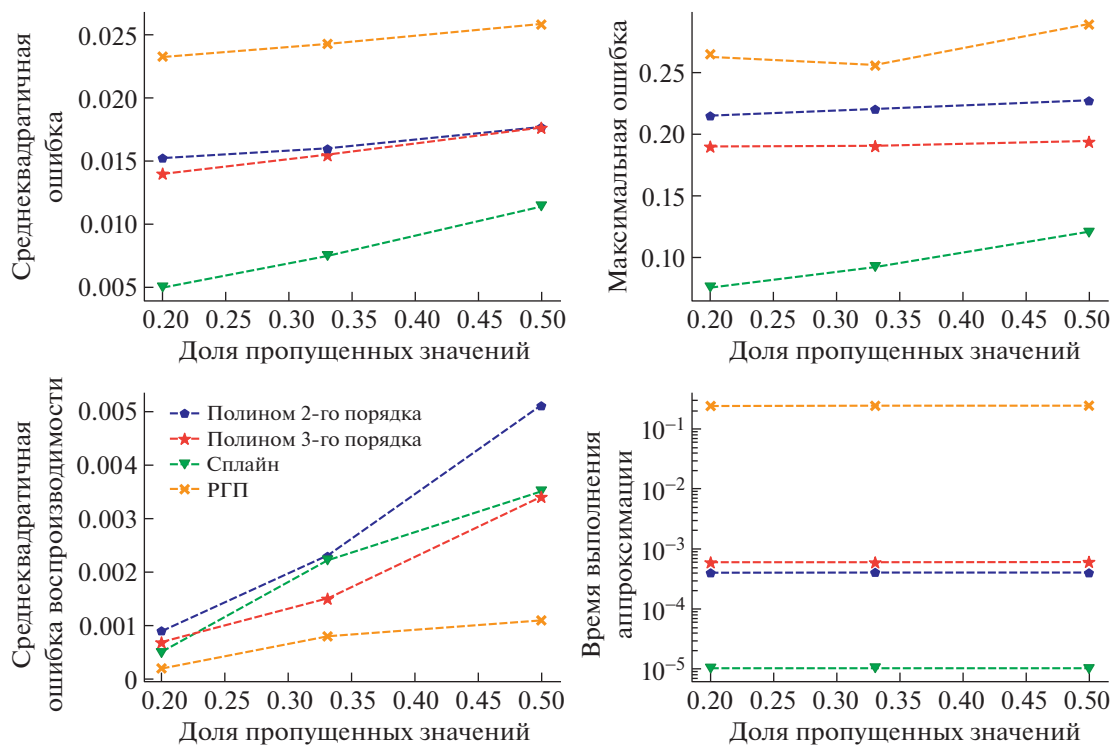
Классификация выполнялась алгоритмом kNN (значение  $k$  выбрано равным пяти) на наборе аппроксимированных временных рядов вегетационного индекса. Тестовая выборка включала в себя по одному полю каждого класса. Остальные поля использовались для обучения классификатора.

По итогам тестирования строились матрицы различий классов и отчеты по классификации, содержащие следующие метрики для каждого класса и взвешенное среднее:

- *precision* (точность) показывает долю объектов, названных классификатором положительными и при этом действительно являющимися положительными, и рассчитывается для каждого класса по формуле:

$$\text{precision} = \frac{TP}{TP + FP},$$

где  $TP$  (True Positive) — количество верно классифицированных положительных примеров,  $FP$  (False Positive) — количество неверно классифицированных положительных примеров.



**Рис. 2.** Зависимости показателей качества от доли пропущенных значений. РГП — регрессия на основе гауссовских процессов.

• *recall* (полнота) показывает, какую долю объектов положительного класса из всех объектов положительного класса нашел алгоритм, и рассчитывается по формуле:

$$recall = \frac{TP}{TP + FP},$$

где *FN* (False Negative) — количество неверно классифицированных отрицательных примеров.

• *F1-score* (F-мера) — среднее гармоническое *precision* и *recall*, рассчитываемое по формуле:

$$F_1 = 2 \frac{precision * recall}{precision + recall}$$

## РЕЗУЛЬТАТЫ

### Сравнение функций аппроксимации

Полученные результаты по аппроксимации временных рядов показывают (рис. 2), что сплайн имеет наименьшие среднеквадратичную и максимальную ошибки, а оптимизация параметров сплайна занимает минимальное время.

В то же время по среднеквадратичной ошибке воспроизводимости сплайн показывает средний результат. Таким образом, из всех функций, предлагаемых для аппроксимации временных рядов вегетационного индекса, по совокупности

выбранных критериев качества наилучшей можно считать сплайн.

### Сравнение методов понижения размерности

Таблицы 1–3 содержат результаты оценок качества для предложенного метода и его модификации с использованием PCA на этапе снижения размерности временных рядов.

Из таблиц видно, что признаки, извлеченные PCA, хоть и повышают качество классификации отдельных классов (строки “Соя” в табл. 1 и 2), но не улучшают среднее качество классификации. В то же время признаки UMAP увеличивают общее качество классификации приблизительно в 1.5 раза.

**Таблица 1.** Отчет по классификации аппроксимированных данных

| Наименование класса | Precision | Recall | F1-Score |
|---------------------|-----------|--------|----------|
| Кукуруза            | 0.51      | 0.34   | 0.41     |
| Подсолнечник        | 0.77      | 0.99   | 0.87     |
| Пшеница             | 0.99      | 0.97   | 0.98     |
| Соя                 | 0.03      | 0.05   | 0.04     |
| Другое              | 0.98      | 0.89   | 0.94     |
| Взвешенное среднее  | 0.66      | 0.65   | 0.66     |

**Таблица 2.** Отчет по классификации с предварительным снижением размерности алгоритмом PCA

| Наименование класса | Precision | Recall | F1-Score |
|---------------------|-----------|--------|----------|
| Кукуруза            | 0.37      | 0.39   | 0.38     |
| Подсолнечник        | 0.52      | 0.58   | 0.55     |
| Пшеница             | 0.99      | 0.99   | 0.99     |
| Соя                 | 0.27      | 0.22   | 0.24     |
| Другое              | 0.98      | 0.89   | 0.93     |
| Взвешенное среднее  | 0.65      | 0.64   | 0.64     |

**Таблица 3.** Отчет по классификации с предварительным снижением размерности алгоритмом UMAP

| Наименование класса | Precision | Recall | F1-Score |
|---------------------|-----------|--------|----------|
| Кукуруза            | 0.96      | 0.99   | 0.98     |
| Подсолнечник        | 0.92      | 1.00   | 0.96     |
| Пшеница             | 0.99      | 0.99   | 0.99     |
| Соя                 | 0.98      | 0.98   | 0.98     |
| Другое              | 0.99      | 0.90   | 0.94     |
| Взвешенное среднее  | 0.97      | 0.96   | 0.96     |

На рис. 3 и 4 представлены результаты классификации всего фрагмента в виде цветного изображения, где цвет каждого пикселя определяется меткой присвоенного ему класса, наложенного на снимок исследуемой местности.

Цветным контуром были выделены соответствующие поля из разметки. Большая часть пикселей изображения имеет красновато-коричневый оттенок, так как были отнесены классификатором к категории “Другое”. Как видно на изображениях, цвета пикселей внутри выделенных контуров более однородны и лучше соответствуют разметке при классификации после предварительного снижения размерности алгоритмом UMAP, чем алгоритмом PCA.

Из табл. 1 и 2 можно видеть, что модификация предложенного алгоритма с использованием PCA вместо UMAP приводит к улучшению качества классификации по F1 показателю только двух классов (“Пшеница” и “Соя”), а по взвешенному среднему проигрывает по всем метрикам. Разница во взвешенных средних показателях F1 для предложенного метода и его модификации отличается в 1.5 раза в пользу предложенного: 0.66 (модификация с PCA) и 0.94 (предложенный).

## ЗАКЛЮЧЕНИЕ

В данной работе рассмотрена задача классификации сельскохозяйственных культур по динамике индекса вегетации NDVI. Акцент в данной работе приходится не на сам алгоритм классификации, в качестве которого используется классический и все еще актуальный метод k-ближайших соседей, а на две вспомогательных подзадачи. Во-первых, это аппроксимация пропущенных из-за облачности значений временного ряда. Во-вто-

**Рис. 3.** Результаты классификации с предварительным снижением размерности алгоритмом PCA.



Рис. 4. Результаты классификации с предварительным снижением размерности алгоритмом UMAP.

рых, это предварительное извлечение признаков из исходного пространства. Проведено исследование моделей аппроксимации временных рядов индекса NDVI. Известная в литературе своим высоким качеством модель на основе сплайнов третьего порядка была впервые соотнесена с другой перспективной моделью на основе гауссовских процессов. В ходе экспериментов было выявлено, что сплайн аппроксимирует исследуемые данные лучше, чем полиномиальные модели и регрессия на основе гауссовских процессов. Предложен алгоритм классификации культур со сплайновой аппроксимацией временных рядов и последующим снижением размерности алгоритмом UMAP. Показано, что предложенный алгоритм классификации по показателю F1 в среднем в 1.5 раза лучше, как его модификации, использующей для извлечения признаков PCA вместо UMAP, так и алгоритма классификации на данных исходной размерности.

#### ИСТОЧНИК ФИНАНСИРОВАНИЯ

Исследование выполнено за счет гранта Российского научного фонда (проект № 20-61-47089).

#### КОНФЛИКТ ИНТЕРЕСОВ

Авторы данной статьи подтвердили отсутствие конфликта интересов, о котором необходимо сообщить.

#### СПИСОК ЛИТЕРАТУРЫ

- Барталев С.А., Лупян Е.А., Нейштадт И.А., Савин И.Ю. Классификация некоторых типов сельскохозяйственных посевов в южных регионах России по спутниковым данным MODIS. *Исследование Земли из космоса*. 2006. Т. 3. С. 68–75.
- Бахтадзе Н.Н., Максимов Е.М., Максимова Н.Е., Дончан Д.М., Кузнецов Д.С., Захаров Э.А. Системы интеллектуального менеджмента для цифрового земледелия. Часть 1. *Информационные технологии и вычислительные системы*. 2020. Т. 2. С. 99–111. <https://doi.org/10.14357/20718632200208>
- Блохина С.Ю. Применение дистанционного зондирования в точном земледелии. *Вестник Российской сельскохозяйственной науки*. 2018. № 5. С. 10–16. <https://doi.org/10.30850/vrtn/2018/5/10-16>
- Бочаров Д.А., Николаев Д.П., Павлова М.А., Тимофеев В.А. Алгоритм детекции и компенсации теней от облаков на мультиспектральных спутниковых снимках для местностей сельскохозяйственных угодий. *Информационные процессы*. 2021. Т. 21. № 4. С. 295–312. [https://doi.org/10.53921/18195822\\_2021\\_21\\_4\\_295](https://doi.org/10.53921/18195822_2021_21_4_295)
- Воробьева Н.С., Чернов А.В. Аппроксимация временных рядов NDVI в задаче раннего распознавания видов сельскохозяйственных культур по космическим снимкам. *Сборник трудов III международной конференции и молодежной школы “Информационные технологии и нанотехнологии” (ИТНТ-2017)- Самара: Новая техника*. Самара. 2017. С. 390–399.
- Павлова М.А., Сидорчук Д.С., Кушев Д.О., Бочаров Д.А., Николаев Д.П. Эквализация условий съемки на основе спектральных моделей для нужд точного земледелия с использованием БПЛА. *Информаци-*

- онные процессы. 2022. Т. 22. № 4. С. 404–413.  
[https://doi.org/10.53921/18195822\\_2022\\_22\\_4\\_404](https://doi.org/10.53921/18195822_2022_22_4_404)
- Плотников Д.Е., Барталев С.А., Жарко В.О., Михайлов В.В., Просяникова О.И. Экспериментальная оценка распознаваемости агрокультур по данным сезонных спутниковых измерений спектральной яркости. *Современные проблемы дистанционного зондирования Земли из космоса*. 2011. Т. 8. № 1. С. 199–208.
- Пугачева И.Ю., Шевырнов А.П. Изучение динамики NDVI посевов сельскохозяйственных культур на территории Красноярского края и Республики Хакасия. *Современные проблемы дистанционного зондирования Земли из космоса*. 2008. Т. 5. № 2. С. 347–351.
- Фирсов Н.А., Подлипнов В.В., Николаев П.П., Машков С.В., Ишкин П.А., Скиданов Р.В., Никоноров А.В. Нейросетевая классификация гиперспектральных изображений растительности с формированием обучающей выборки на основе адаптивного вегетационного индекса. *Компьютерная оптика*. 2021. Т. 45. № 6. С. 887–896.  
<https://doi.org/10.18287/2412-6179-CO-1038>
- Черепанов А.С., Дружинина Е.Г. Спектральные свойства растительности и вегетационные индексы. *Геоматика*. 2009. № 3. С. 28–32.
- Якушев В.П., Дубенок Н.Н., Лупян Е.А. Опыт применения и перспективы развития технологий дистанционного зондирования Земли для сельского хозяйства. *Современные проблемы дистанционного зондирования Земли из космоса*. 2019. Т. 16. № 3. С. 11.
- Abe B.T., Jordaаn J.A. Hyperspectral image classification based on NMF Features Selection Method. *Sixth International Conference on Machine Vision (ICMV 2013)*. SPIE, 2013. Т. 9067. С. 114–119.  
<https://doi.org/10.1117/12.2050072>
- Belda S., Pipia L., Morcillo-Pallarés P., Verrelst J. Optimizing gaussian process regression for image time series gap-filling and crop monitoring. *Agronomy*. 2020. Т. 10. № 5. С. 618.  
<https://doi.org/10.3390/agronomy10050618>
- Bouteldja S., Kourgli A. A comparative analysis of SVM, K-NN, and decision trees for high resolution satellite image scene classification. *Twelfth International Conference on Machine Vision (ICMV 2019)*. SPIE, 2020. Т. 11433. С. 410–416.  
<https://doi.org/10.1117/12.2557563>
- Chakhar A., Hernández-López D., Ballesteros R., Moreno M.A. Improving the accuracy of multiple algorithms for crop classification by integrating sentinel-1 observations with sentinel-2 data. *Remote Sensing*. 2021. Т. 13. № 2. С. 243.  
<https://doi.org/10.3390/rs13020243>
- Gilbertson J.K., Van Niekerk A. Value of dimensionality reduction for crop differentiation with multi-temporal imagery and machine learning. *Computers and Electronics in Agriculture*. 2017. Т. 142. С. 50–58.  
<https://doi.org/10.1016/j.compag.2017.08.024>
- Groten S.M.E. NDVI—crop monitoring and early yield assessment of Burkina Faso. *International Journal of Remote Sensing*. 1993. Т. 14. № 8. С. 1495–1515.  
<https://doi.org/10.1080/01431169308953983>
- Hird J.N., McDermid G.J. Noise reduction of NDVI time series: An empirical comparison of selected techniques. *Remote Sensing of Environment*. 2009. Т. 113. № 1. С. 248–258.  
<https://doi.org/10.1016/j.rse.2008.09.003>
- Li J., Shen Y., Yang C. An adversarial generative network for crop classification from remote sensing timeseries images. *Remote Sensing*. 2020. Т. 13. № 1. С. 65.  
<https://doi.org/10.3390/rs13010065>
- McInnes L., Healy J., Melville J. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*. 2018.  
<https://doi.org/10.48550/arXiv.1802.03426>
- Murmu S., Biswas S. Application of fuzzy logic and neural network in crop classification: a review. *Aquatic Procedia*. 2015. Т. 4. С. 1203–1210.  
<https://doi.org/10.1016/j.aqpro.2015.02.153>
- Orynbaikyzy A., Gessner U., Conrad C. Crop type classification using a combination of optical and radar remote sensing data: A review. *International journal of remote sensing*. 2019. Т. 40. № 17. С. 6553–6595.  
<https://doi.org/10.1080/01431161.2019.1569791>
- Reedha R., Dericquebourg E., Canals R., Hafiane A. Transformer neural network for weed and crop classification of high resolution UAV images. *Remote Sensing*. 2022. Т. 14. № 3. С. 592.  
<https://doi.org/10.3390/rs14030592>
- Rußwurm M., Körner M. Self-attention for raw optical satellite time series classification. *ISPRS journal of photogrammetry and remote sensing*. 2020. Т. 169. С. 421–435.  
<https://doi.org/10.1016/j.isprsjprs.2020.06.006>
- Sishodia R.P., Ray R.L., Singh S.K. Applications of remote sensing in precision agriculture: A review. *Remote Sensing*. 2020. Т. 12. № 19. С. 3136.  
<https://doi.org/10.3390/rs12193136>
- Sun R., Chen S., Su H., Mi C., Jin N. The effect of NDVI time series density derived from spatiotemporal fusion of multisource remote sensing data on crop classification accuracy. *ISPRS International Journal of Geo-Information*. 2019. Т. 8. № 11. С. 502.  
<https://doi.org/10.3390/ijgi8110502>
- Velliangiri S., Alagumuthukrishnan S., Thankumar S.I. A review of dimensionality reduction techniques for efficient computation. *Procedia Computer Science*. 2019. Т. 165. С. 104–111.  
<https://doi.org/10.1016/j.procs.2020.01.079>
- Yang S., Gu L., Li X., Jiang T., Ren R. Crop classification method based on optimal feature selection and hybrid CNN-RF networks for multi-temporal remote sensing imagery. *Remote sensing*. 2020. Т. 12. № 19. С. 3119.  
<https://doi.org/10.3390/rs12193119>
- Zhang S., Lei Y., Wang L., Li H., Zhao H. Crop classification using MODIS NDVI data denoised by wavelet: A case study in Hebei Plain, China. *Chinese Geographical Science*. 2011. Т. 21. С. 322–333.  
<https://doi.org/10.1007/s11769-011-0472-2>

## Classification of crops by NDVI time series of reduced dimensionality

M. A. Pavlova<sup>a, #</sup>, D. S. Sidorchuk<sup>a</sup>, and D. A. Bocharov<sup>a</sup>

<sup>a</sup>*Institute for Information Transmission Problems. A.A. Kharkevich RAS  
127051, Moscow, Bolshoi Karetny lane, 19, Russia*

<sup>#</sup>*E-mail: pavlovamar96@gmail.com*

The paper considers the problem of classification of agricultural crops. As is known, to solve this problem, it is much more efficient to use not instantaneous remote sensing data or calculated vegetation indices, but their historical series. Time series formed by index values for a fixed spatial point at different dates are characterized by a high level of missing values, caused primarily by cloudiness on some dates. A study of known methods of time series approximation has been carried out. The question of whether reducing the dimensionality of the approximated time series can improve the quality of crops classification is also investigated. In the experimental part of the work, NDVI time series calculated from the Sentinel-2 multispectral satellite data were used. The classification of corn, sunflower, wheat and soybeans was studied. The paper shows that UMAP usage for dimensionality reduction leads to 1.5 times increase of classification quality in terms of average the F1-measure compared to using the original dimension data. A new crop classification method based on cubic spline approximation of NDVI time series, extraction of features of low dimension by the UMAP algorithm and their classification by the k nearest neighbors method is proposed.

**Keywords:** Remote sensing, crop classification, time series, NDVI, time series fitting, feature extraction, dimensionality reduction, UMAP

### REFERENCES

- Bartalev S.A., Lupyan E.A., Neishtadt I.A., Savin I.Yu. Klassifikatsiya nekotorykh tipov sel'skokhozyaistvennykh posevov v yuzhnykh regionakh Rossii po sputnikovym dannym MODIS [Classification of some types of agricultural crops in the southern regions of Russia according to MODIS satellite data.]. *Issledovanie Zemli iz kosmosa* [Earth exploration from space]. 2006. V. 3. P. 68–75 (in Russian).
- Bakhtadzel N., Maximov E., Maximova N., Donchan D., Kuznetsov D., Zakharov E. Intelligent Management Systems for Digital Farming. Part 1. Informacionnye tekhnologii i vychislitel'nye sistemy [Information technologies and computing systems]. 2020. V. 2. P. 99–111. <https://doi.org/10.14357/20718632200208>
- Blokhina S.Yu. The application of remote sensing in precision agriculture. *Vestnik of the Russian agricultural science*. 2018. (5). P. 10–16 (in Russian). <https://doi.org/10.30850/vrsn/2018/5/10-16>
- Bocharov D.A., Nikolaev D.P., Pavlova M.A., Timofeev V.A. Cloud Shadows Detection and Compensation Algorithm on Multispectral Satellite Images for Agricultural Regions. *JCTE*. 2022. V. 67. № 6. P. 728–739. <https://doi.org/10.1134/S1064226922060171>
- Vorob'eva N.S., Chernov A.V. Approksimatsiya vremennykh ryadov NDVI v zadache rannego raspoznavaniya vidov sel'skokhozyaistvennykh kul'tur po kosmicheskim snimkam [Approximation of NDVI time series in the problem of early recognition of crop species from satellite images]. *Sbornik trudov III mezhdunarodnoi konferentsii i molodezhnoi shkoly "Informatsionnye tekhnologii i nanotekhnologii" (ITNT-2017)-Samara: Novaya tekhnika* [Proceedings of the III International Conference and Youth School "Information Technology and Nanotechnology" (ITNT-2017) – Samara: New technology]. Samara. 2017. P. 390–399 (in Russian).
- Pavlova M.A., Sidorchuk D.S., Kushchev D.O., Bocharov D.A., Nikolaev D.P. Equalization of Shooting Conditions Based on Spectral Models for the Needs of Precision Agriculture Using UAVs. *JCTE*. 2022. V. 67. № 2. <https://doi.org/10.1134/S1064226922140066>
- Plotnikov D.E., Bartalev S.A., Zharko V.O., Mihailov V.V., Prosyannikova O.I. An experimental assessment of crop types recognisability using time-series of intra-seasonal spectral reflectance measurements by satellite sensor. *Sovremennye problemy distantsionnogo zondirovaniya Zemli iz kosmosa* [Modern problems of remote sensing of the Earth from space]. 2011. V. 8 (1). P. 199–208 (in Russian).
- Pugacheva I.Yu., Shevyrnogov A.P. Izuchenie dinamiki NDVI posevov sel'skokhozyaistvennykh kul'tur na territorii Krasnoyarskogo kraia i Respubliki Khakasiya [The study of the dynamics of NDVI crops of agricultural crops in the territory of the Krasnoyarsk Territory and the Republic of Khakassia]. *Sovremennye problemy distantsionnogo zondirovaniya Zemli iz kosmosa* [Modern problems of remote sensing of the Earth from space]. 2008. V. 5 (2). P. 347–351 (in Russian).
- Firsov N., Podlipnov V., Ivliev N., Nikolaev. P, Mashkov S., Ishkin P., Skidanov R., Nikonorov A. Neural network-aided classification of hyperspectral vegetation images with a training sample generated using an adaptive vegetation index. *Computer Optics*. 2021. V. 45 (6). P. 887–896 (in Russian). <https://doi.org/10.18287/2412-6179-CO-1038>
- Cherepanov A.S., Druzhinina E.G. Spektral'nye svoystva rastitel'nosti i vegetatsionnye indeksy [Spectral properties of vegetation and vegetation indices]. *Geomatika* [Geomatics]. 2009. V. 3. P. 28–32 (in Russian).
- Yakushev V.P., Dubenok N.N., Loupian E.A. Earth remote sensing technologies for agriculture: application experience and development prospects. *Sovremennye problemy distantsionnogo zondirovaniya Zemli iz kosmosa* [Modern problems of remote sensing of the Earth from space]. 2019. V. 16 (3). P. 11 (in Russian).

- Abe B.T., Jordaan J.A. Hyperspectral image classification based on NMF Features Selection Method. *Sixth International Conference on Machine Vision (ICMV 2013)*. SPIE, 2013. V. 9067. P. 114–119. <https://doi.org/10.1117/12.2050072>
- Belda S., Pipia L., Morcillo-Pallarés P., Verrelst J. Optimizing gaussian process regression for image time series gap-filling and crop monitoring. *Agronomy*. 2020. V. 10 (5). P. 618. <https://doi.org/10.3390/agronomy10050618>
- Bouteldja S., Kourgli A. A comparative analysis of SVM, K-NN, and decision trees for high resolution satellite image scene classification. *Twelfth International Conference on Machine Vision (ICMV 2019)*. SPIE, 2020. V. 11433. P. 410–416. <https://doi.org/10.1117/12.2557563>
- Chakhar A., Hernández-López D., Ballesteros R., Moreno M.A. Improving the accuracy of multiple algorithms for crop classification by integrating sentinel-1 observations with sentinel-2 data. *Remote Sensing*. 2021. V. 13 (2). P. 243. <https://doi.org/10.3390/rs13020243>
- Gilbertson J.K., Van Niekerk A. Value of dimensionality reduction for crop differentiation with multi-temporal imagery and machine learning. *Computers and Electronics in Agriculture*. 2017. V. 142. P. 50–58. <https://doi.org/10.1016/j.compag.2017.08.024>
- Groten S.M.E. NDVI—crop monitoring and early yield assessment of Burkina Faso. *International Journal of Remote Sensing*. 1993. V. 14 (8). P. 1495–1515. <https://doi.org/10.1080/01431169308953983>
- Hird J.N., McDermid G.J. Noise reduction of NDVI time series: An empirical comparison of selected techniques. *Remote Sensing of Environment*. 2009. V. 113 (1). P. 248–258. <https://doi.org/10.1016/j.rse.2008.09.003>
- Li J., Shen Y., Yang C. An adversarial generative network for crop classification from remote sensing timeseries images. *Remote Sensing*. 2020. V. 13 (1). P. 65. <https://doi.org/10.3390/rs13010065>
- McInnes L., Healy J., Melville J. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*. 2018. <https://doi.org/10.48550/arXiv.1802.03426>
- Murmu S., Biswas S. Application of fuzzy logic and neural network in crop classification: a review. *Aquatic Procedia*. 2015. V. 4. P. 1203–1210. <https://doi.org/10.1016/j.aqpro.2015.02.153>
- Orynbaikyzy A., Gessner U., Conrad C. Crop type classification using a combination of optical and radar remote sensing data: A review. *International journal of remote sensing*. 2019. V. 40 (17). P. 6553–6595. <https://doi.org/10.1080/01431161.2019.1569791>
- Reedha R., Dericquebourg E., Canals R., Hafiane A. Transformer neural network for weed and crop classification of high resolution UAV images. *Remote Sensing*. 2022. V. 14 (3). P. 592. <https://doi.org/10.3390/rs14030592>
- Rußwurm M., Körner M. Self-attention for raw optical satellite time series classification. *ISPRS journal of photogrammetry and remote sensing*. 2020. V. 169. P. 421–435. <https://doi.org/10.1016/j.isprsjprs.2020.06.006>
- Sishodia R.P., Ray R.L., Singh S.K. Applications of remote sensing in precision agriculture: A review. *Remote Sensing*. 2020. V. 12 (19). P. 3136. <https://doi.org/10.3390/rs12193136>
- Sun R., Chen S., Su H., Mi C., Jin N. The effect of NDVI time series density derived from spatiotemporal fusion of multisource remote sensing data on crop classification accuracy. *ISPRS International Journal of Geo-Information*. 2019. V. 8 (11). P. 502. <https://doi.org/10.3390/ijgi8110502>
- Velliangiri S., Alagumuthukrishnan S., Thankumar S.I. A review of dimensionality reduction techniques for efficient computation. *Procedia Computer Science*. 2019. V. 165. P. 104–111. <https://doi.org/10.1016/j.procs.2020.01.079>
- Yang S., Gu L., Li X., Jiang T., Ren R. Crop classification method based on optimal feature selection and hybrid CNN-RF networks for multi-temporal remote sensing imagery. *Remote sensing*. 2020. V. 12 (19). P. 3119. <https://doi.org/10.3390/rs12193119>
- Zhang S., Lei Y., Wang L., Li H., Zhao H. Crop classification using MODIS NDVI data denoised by wavelet: A case study in Hebei Plain, China. *Chinese Geographical Science*. 2011. V. 21. P. 322–333. <https://doi.org/10.1007/s11769-011-0472-2>