

УДК: 004.932

Оптимизация размера классификатора при сегментации трехмерных точечных образов древесной растительности

И. М. Никольский

Московский государственный университет им. М. В. Ломоносова,
Россия, 119991, г. Москва, Воробьевы горы, д. 1

E-mail: inikolsky@cs.msu.ru

Получено 10.05.2025, после доработки — 28.07.2025.

Принято к публикации 30.07.2025.

Появление технологий лазерного сканирования произвело настоящую революцию в лесном хозяйстве. Их использование позволило перейти от изучения лесных массивов с помощью ручных измерений к компьютерному анализу точечных стереоизображений, называемых облаками точек.

Автоматическое вычисление некоторых параметров деревьев (таких как диаметр ствола) по облаку точек требует удаления точек листвы. Для выполнения этой операции необходима предварительная сегментация стереоизображения на классы «листва» и «ствол». Решение этой задачи зачастую включает использование методов машинного обучения.

Одним из самых популярных классификаторов, используемых для сегментации стереоизображений деревьев, является случайный лес. Этот классификатор достаточно требователен к объему памяти. В то же время размер модели машинного обучения может быть критичным при необходимости ее пересылки, что требуется, например, при выполнении распределенного обучения. В данной работе ставится цель найти классификатор, который был бы менее требовательным по памяти, но при этом имел бы сравнимую точность сегментации. Поиск выполняется среди таких классификаторов, как логистическая регрессия, наивный байесовский классификатор и решающее дерево. Кроме того, исследуется способ уточнения сегментации, выполненной решающим деревом, с помощью логистической регрессии.

Эксперименты проводились на данных из коллекции университета Гейдельберга. Было показано, что классификация с помощью решающего дерева, корректируемая с помощью логистической регрессии, способна давать результат, лишь немного проигрывающий результату случайного леса по точности, затрачивая при этом меньше времени и оперативной памяти. Разница в сбалансированной точности составляет не более процента на всех рассмотренных облаках, при этом суммарный размер и время предсказания классификаторов решающего дерева и логистической регрессии на порядок меньше, чем у случайного леса.

Ключевые слова: лазерное сканирование, облако точек, машинное обучение, сегментация

UDC: 004.932

Classifier size optimisation in segmentation of three-dimensional point images of wood vegetation

I. M. Nikolsky

Lomonosov Moscow State Institute University,
1 Vorobyovy Gory, Moscow, 119991, Russia

E-mail: inikolsky@cs.msu.ru

Received 10.05.2025, after completion – 28.07.2025.

Accepted for publication 30.07.2025.

The advent of laser scanning technologies has revolutionized forestry. Their use made it possible to switch from studying woodlands using manual measurements to computer analysis of stereo point images called point clouds.

Automatic calculation of some tree parameters (such as trunk diameter) using a point cloud requires the removal of foliage points. To perform this operation, a preliminary segmentation of the stereo image into the “foliage” and “trunk” classes is required. The solution to this problem often involves the use of machine learning methods.

One of the most popular classifiers used for segmentation of stereo images of trees is a random forest. This classifier is quite demanding on the amount of memory. At the same time, the size of the machine learning model can be critical if it needs to be sent by wire, which is required, for example, when performing distributed learning. In this paper, the goal is to find a classifier that would be less demanding in terms of memory, but at the same time would have comparable segmentation accuracy. The search is performed among classifiers such as logistic regression, naive Bayes classifier, and decision tree. In addition, a method for segmentation refinement performed by a decision tree using logistic regression is being investigated.

The experiments were conducted on data from the collection of the University of Heidelberg. The collection contains hand-marked stereo images of trees of various species, both coniferous and deciduous, typical of the forests of Central Europe.

It has been shown that classification using a decision tree, adjusted using logistic regression, is able to produce a result that is only slightly inferior to the result of a random forest in accuracy, while spending less time and RAM. The difference in balanced accuracy is no more than one percent on all the clouds considered, while the total size and inference time of the decision tree and logistic regression classifiers is an order of magnitude smaller than of the random forest classifier.

Keywords: laser scanning, point cloud, machine learning, segmentation

Citation: *Computer Research and Modeling*, 2025, vol. 17, no. 4, pp. 665–675 (Russian).

Введение

Технологии лазерного сканирования позволяют получать трехмерный цифровой образ предмета либо сцены, состоящий из точек с координатами (x, y, z) . Стереοизображения такого типа принято называть *облаками точек*. Современные сканеры позволяют получать очень подробное структурное представление объекта, генерируя сотни тысяч и даже миллионы точек в секунду. Благодаря этому, а также постепенному снижению стоимости сканирующего оборудования технология лазерного сканирования проникает во все новые сферы человеческой деятельности, такие как мониторинг путей сообщения [Hu et al., 2021], сохранение культурного наследования [Alshawabkeh et al., 2020] и др.

Одним из примеров успешного использования рассматриваемой технологии является лесное хозяйство [Нууррә et al., 2012]. Лазерное сканирование участков леса, выполняемое с воздуха либо с земли, позволяет заменить ручные измерения компьютерным анализом трехмерных снимков с целью определения различных параметров, имеющих значение для учета лесных массивов: количество деревьев, их породный состав и т. д. (см., например, [Liang, 2016]).

Анализ отдельных деревьев также представляет интерес для лесного хозяйства (оценка высоты, толщины ствола, объема зеленой массы и других характеристик [Park, Lee, 2023]), а также для автоматизированного сбора урожая с применением роботов [Qiu et al., 2025]. В этом случае необходимо отделение кроны от ствола точек, что требует выполнения предварительной *сегментации* стереοизображения дерева, т. е. отнесения каждой точки облака к одному из двух классов. Для решения данной задачи с успехом применяются методы машинного обучения, как классические (случайный лес, метод опорных векторов и др.), так и нейросетевые.

Задача сегментации стереοизображений деревьев имеет особые характеристики. Изображения достаточно велики (состоят из нескольких сот тысяч и даже миллионов точек), и, как правило, имеет место дисбаланс между точками различных классов, причем в одних изображениях доминирует листва, в других — ствол. В связи с этим для оценки качества сегментации целесообразно использовать метрику «сбалансированная точность» (balanced accuracy), которая часто применяется в подобных случаях.

Одним из наиболее популярных классификаторов, используемых для сегментации облаков точек, является случайный лес. К сожалению, данный классификатор имеет достаточно большие требования к объему памяти, что особенно заметно при работе с большими облаками точек. Фактор размера модели машинного обучения играет важную роль при необходимости ее пересылки в случае федеративного и других разновидностей распределенного обучения.

Данная работа посвящена поиску классификатора, который был бы более экономичным, чем случайный лес, в смысле потребляемой памяти, но при этом обладал бы сравнимой сбалансированной точностью.

В качестве кандидатов на роль такого более экономичного классификатора нами рассматривались логистическая регрессия, наивный байесовский классификатор и решающее дерево. Отметим, что, хотя нет гарантий выполнения условий, необходимых для применения наивного байесовского классификатора, тем не менее, как показано в [Domingos, Pazzani, 1997], это не мешает ему показывать хорошие результаты во многих практических случаях.

Кроме того, были проведены эксперименты с уточнением сегментации, выполненной некоторым классификатором, каким-либо другим классификатором. В ходе такого уточнения точки, распознанные первым классификатором как принадлежащие одному из классов, подаются на вход второму классификатору. Именно такой комбинированный способ сегментации показал наилучшие результаты.

Сравнение экономических классификаторов выполнялось с классификатором «случайный лес» с количеством деревьев 10, 20, 30. Большее количество деревьев в случайном лесе не рассматривалось, так как расчеты показали незначительное влияние количества деревьев на сбалансированную точность (к подобным выводам пришли и авторы работы [Lin et al., 2021]).

Обзор литературы

Применению методов машинного обучения для сегментации облаков точек посвящена обширная литература. Так, например, в работе [Lin et al., 2021] рассматривается сегментация стереоизображений деревьев на классы «ствол» и «ветви/листва». Положительным считается первый класс. В качестве классификатора используется случайный лес из 30 решающих деревьев. Авторы отмечают, что увеличение количества деревьев не приводит к существенному улучшению качества сегментации. Исследования проводились на наборе трехмерных изображений лесных деревьев, состоящем только из сосен *Pinus tabulaeformis*. Вектор признаков состоял из 8 признаков, включавших высоту (z -координату) точки, локальную плотность и некоторые геометрические характеристики, вычисляемые с помощью метода главных компонент. Было показано, что наиболее значимыми являются высота и плотность. Еще одним примером использования случайного леса для сегментации облаков точек является работа [Alfio et al., 2024], где данный алгоритм используется для сегментации трехмерных изображений городских ландшафтов.

Работа [Yun et al., 2016] также посвящена сегментации стереоизображений деревьев, однако на этот раз использовался метод опорных векторов. Набор признаков несколько отличается от того, что использовался в [Lin et al., 2021]. Эксперименты проводились на коллекции точечных стереоизображений широколиственных экзотических деревьев различных пород (гевеи, михелии, сакуры и т. д.).

Глубокое обучение также широко используется в задачах распознавания и сегментации облаков точек. Наиболее известными архитектурами являются PointNet [Qi et al., 2016] и PointNet++ [Qi et al., 2017]. Подробный обзор использования методов глубокого обучения при сегментации трехмерных изображений деревьев можно найти в [Sarker et al., 2023].

Большой интерес представляет использование последовательности алгоритмов сегментации, в которой каждый следующий корректирует результаты классификации точек, полученные предыдущим. Так, например, в работе [Sun et al., 2021] описан алгоритм, классифицирующий точки в три этапа. На первом этапе точки классифицируются на основе интенсивности отраженного лазерного луча, затем выполняется уточнение на основе показателя локальной плотности облака, после чего окончательное разбиение на классы получается с помощью вокселизации. В работе [Liu et al., 2021] предлагается другой способ, который сначала выполняет сегментацию стереоизображения дерева с помощью геометрических признаков, после чего результат уточняется с помощью алгоритма роста региона (region-growth) и стратификации точек по высоте.

Постановка эксперимента

Данные

Вычислительные эксперименты проводились на стереоизображениях (облаках точек) из коллекции Университета Гейдельберга [Weiser et al., 2024] (см. табл. 1). Данные были получены с помощью лазерного сканирования в лесах Бреттена и Карлсруэ в земле Баден-Вюртемберг на юго-западе Германии. Стереоизображения в коллекции размечены, т. е. каждая точка имеет метку «ствол» либо «листва». Положительным является класс «листва». Физически данные хранятся в файлах формата laz.

Таблица 1. Размер тестовых стереоизображений

Номер облака	Русское название	Латинское название	Тип	Количество точек
1	Бук европейский	Fagus sylvatica	лиственное	2 611 685
2	Дуб скальный	Quercus petraea	лиственное	479 034
3	Ель обыкновенная	Picea abies	хвойное	1 710 699
4	Дуб красный	Quercus rubra	лиственное	2 918 205
5	Дуб красный	Quercus rubra	лиственное	4 839 094
6	Сосна обыкновенная	Pinus sylvestris	хвойное	3 936 054
7	Клен белый	Acer pseudoplatanus	лиственное	6 382 773
8	Сосна обыкновенная	Pinus sylvestris	хвойное	1 322 666
9	Псевдотсуга мензиса	Pseudotsuga menziesii	хвойное	5 852 254

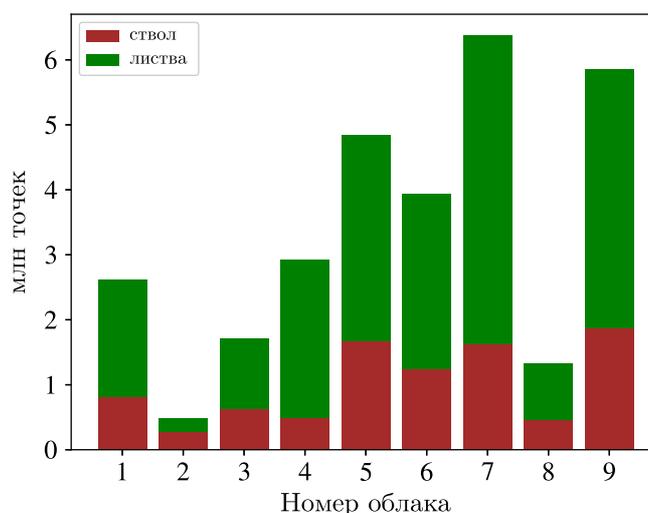


Рис. 1. Количество точек в рассматриваемых стереоизображениях

Рассматриваемые стереоизображения достаточно подробны и состоят из миллионов точек, за исключением одного, включающего порядка полумиллиона точек. Размеры изображений (в миллионах точек) показаны на рис. 1. Отметим, что в отличие от [Lin et al., 2021] в используемых нами изображениях точек листвы намного больше, чем точек ствола, в то время как в данных, рассматриваемых в [Lin et al., 2021], наблюдается обратная ситуация.

Предобработка

Облака точек, участвующие в вычислительных экспериментах, прореживались при загрузке в 10 раз для ускорения векторизации. После загрузки выполнялось масштабирование с помощью MinMaxScaler, которое состоит в применении к каждой из пространственных координат следующего преобразования: $t(s) = \frac{s-s_{\min}}{s_{\max}-s_{\min}}$. Здесь s_{\max} и s_{\min} — максимальное и минимальное значения признака s соответственно. Такое преобразование данных традиционно используется при классификации с помощью решающих деревьев.

Векторы признаков (определены ниже) масштабировались с помощью StandardScaler в соответствии с формулой $f' = \frac{f-\mu}{\sigma}$, где f — исходное значение признака, μ и σ — математическое ожидание и стандартное отклонение данного признака соответственно, f' — стандартизованное значение. Такое масштабирование позволяет нивелировать влияние признаков с большим диапазоном значений.

Признаки

При векторизации каждой точке облака сопоставляется вектор, состоящий из ее пространственных координат (x, y, z) в совокупности с набором геометрических признаков, традиционно используемых при сегментации облаков точек.

Данные признаки определяются следующим образом. Пусть p_i — точка облака P с координатами (x_i, y_i, z_i) . Возьмем k ближайших к p_i точек облака P . Вычислим корреляционную матрицу:

$$\text{Cov}(p_i) = \frac{1}{k} \sum_{j=1}^k (p_j - \bar{p})(p_j - \bar{p})^T,$$

где \bar{p} — центроид k ближайших соседей точки p_i , p_j — j -й сосед точки p_i .

Обозначим упорядоченные по убыванию собственные значения (с. з.) этой матрицы как $\lambda_1, \lambda_2, \lambda_3$, где $\lambda_1 > \lambda_2 > \lambda_3$. Далее определим следующие характеристики окрестности точки p_i :

$$\text{индекс линейности } L_i = \frac{\lambda_1 - \lambda_2}{\lambda_1},$$

$$\text{индекс планарности } P_i = \frac{\lambda_2 - \lambda_3}{\lambda_1},$$

$$\text{индекс разброса точек } S_i = \frac{\lambda_3}{\lambda_1},$$

$$\text{индекс кривизны } C_i = \frac{\lambda_3}{\lambda_1 + \lambda_2 + \lambda_3},$$

$$\text{индекс вертикальности } V_i = \langle \mathbf{e}_3, (0, 0, 1) \rangle,$$

где \mathbf{e}_3 — собственный вектор, отвечающий с. з. λ_3 , $\langle \cdot, \cdot \rangle$ — скалярное произведение. Окончательно каждой точке облака p_i сопоставляется вектор признаков $\mathbf{v}(p_i) = (x_i, y_i, z_i, L_i, P_i, S_i, C_i, V_i)$.

Введенные таким образом геометрические признаки характеризуют окрестность классифицируемой точки. Индексы линейности, планарности и разброса точек (иногда его называют *индексом сферичности*) позволяют классификатору различать линейные, плоские и объемные элементы. Индекс кривизны (или индекс *вариации поверхности*) позволяет идентифицировать геометрические изменения поверхности. Индекс вертикальности, который представляет собой z -компоненту нормали к плоскости наилучшего приближения, показывает отклонение точек окрестности от вертикали. Использование пространственных координат обусловлено тем, что листва, как правило, находится высоко от земли и на удалении от вертикальной оси дерева.

Оценка эффективности

Для каждого облака точек выполнялась тренировка исследуемых классификаторов на обучающей выборке, состоящей из 10 % точек облака. Выбор настолько небольшой части данных в качестве обучающей выборки обусловлен тематикой данной работы, связанной с минимизацией размера классификатора. В работе [Yun et al., 2016] было показано, что 5–10 % точек отсканированного дерева достаточно для эффективного обучения классификатора. Кроме того, с учетом большой плотности каждого из рассматриваемых облаков точек десятипроцентная обучающая выборка многократно перекрывает по количеству точек эмпирики для минимального размера выборки, принятые в классической статистике, где на каждый признак требуется 10/20/50 наблюдений.

В качестве положительного класса использовался класс «листва». Тестирование выполнялось на остальных точках облака. Для каждого классификатора измерялись время обучения, время предсказания, а также размер в байтах, который определялся путем сохранения классификатора с помощью библиотеки pickle.

Для оценки качества сегментации использовалась метрика «сбалансированная точность», которую определим следующим образом. Введем обозначения: r_P — чувствительность (sensitivity), она же полнота (recall) для положительного класса, r_N — избирательность (specificity, полнота для отрицательного класса), где

$$r_P = \frac{TP}{TP + FN},$$
$$r_N = \frac{TN}{TN + FP}.$$

Здесь TP , TN , FN , FP — количество соответственно истинно положительных, истинно отрицательных, ложноположительных и ложноотрицательных результатов классификации. Теперь мы можем определить сбалансированную точность с помощью следующей формулы:

$$a_b = \frac{r_P + r_N}{2}.$$

Классификаторы

Как уже было сказано, в данной работе рассматривались следующие бинарные классификаторы: наивный байесовский классификатор, логистическая регрессия, решающее дерево и случайный лес. Во всех классификаторах никакие специальные настройки для компенсации несбалансированности выборки не использовались.

В качестве наивного байесовского классификатора использовался вариант, в котором предполагается, что признаки следуют нормальному распределению (Gaussian Naive Bayes). Не делалось каких-либо предположений относительно априорного распределения, априорные вероятности оценивались непосредственно из данных. Что касается логистической регрессии, то она использовалась с регуляризацией L2 и алгоритмом оптимизации L-BFGS.

Перейдем к описанию гиперпараметров решающего дерева. Аналогично [Lin et al., 2021] в наших экспериментах не использовались какие-либо меры против переобучения. В частности, не применялась обрезка (прунинг) дерева, максимальная глубина дерева и количество листовых узлов не ограничивались.

Принципы разбиения узлов были следующими. В качестве критерия качества разделения узлов использовался критерий Джини, минимальное количество точек, необходимых для разбиения внутреннего узла, было равно 2. При разбиении узла всегда выполнялся поиск наилучшего разбиения содержащихся в нем точек, при этом рассматривалось все множество признаков.

В экспериментах со случайным лесом гиперпараметры решающих деревьев были такими же.

Программная реализация и параметры аппаратуры

Для проведения вычислительных экспериментов была написана программа на языке Python 3. Реализации классификаторов были взяты из библиотеки sklearn версии 1.3.2. При построении ковариационной матрицы для ускорения поиска соседних точек использовалось kd-дерево, также из sklearn. Загрузка laz-файлов выполнялась с помощью библиотеки laspy. Кроме того, при написании программы использовалась библиотека numru версии 1.24.4.

Вычислительные эксперименты проводились на компьютере со следующими характеристиками: 64-битный процессор AMD EPYC 7763 с частотой 2450 МГц, объем ОЗУ — 1 Гб. Стоит пояснить актуальность выбора машины со столь небольшим объемом ОЗУ. Поскольку

мы ставим перед собой задачу минимизацию размера классификатора, представляет интерес проведение экспериментов в условиях ограниченной оперативной памяти. Отметим, что объем ОЗУ 1 Гб встречается на устройствах в области мобильной робототехники (см. [Wozniak et al., 2025]), а также в некоторых одноплатных компьютерах, используемых при построении роботизированных платформ (Beagle board).

Графический ускоритель в экспериментах не использовался, все вычисления выполнялись на центральном процессоре.

Результаты вычислительных экспериментов

В данном разделе представлены результаты сегментирования облаков точек различными классификаторами. Измерялись время обучения и предсказания, а также сбалансированная точность сегментации.

В табл. 2 представлены размеры исследуемых классификаторов. Как видим, размеры простейших моделей машинного обучения (логистическая регрессия и наивный байесовский классификатор) одинаковы для всех облаков точек и равны 700 и 780 байт соответственно. Все остальные классификаторы (решающее дерево и варианты случайного леса) имеют размер, коррелирующий с размером соответствующего облака. Таким образом, наименьший размер имеет классификатор для облака 2, наибольший — для облака 7. При этом если размер простейших классификаторов находится в пределах килобайта, то размер решающих деревьев (за исключением самого маленького для облака 2) составляет уже сотни килобайт. Размер классификаторов «случайный лес» измеряется мегабайтами, довольно быстро возрастая при увеличении количества деревьев в лесе. Для случайного леса с 30 деревьями размеры самых больших классификаторов (для облаков 4, 5, 6, 7) уже составляет 10–20 мегабайт.

Таблица 2. Размер классификаторов, байт

Облако	Логистическая регрессия	Наивный Байес	Решающее дерево	Сл. лес (10)	Сл. лес (20)	Сл. лес (30)
Облако 1	700	780	234 637	2 310 262	4 591 692	6 889 282
Облако 2	700	780	48 388	583 346	1 156 580	1 726 783
Облако 3	700	780	165 517	1 719 542	3 440 172	5 118 722
Облако 4	700	780	285 837	2 611 382	5 210 252	7 840 802
Облако 5	700	780	466 006	5 294 672	10 593 312	15 865 872
Облако 6	700	780	317 837	3 487 943	6 897 454	10 453 356
Облако 7	700	780	838 166	7 888 752	15 681 312	23 480 272
Облако 8	700	780	131 917	1 413 302	2 821 292	4 195 362
Облако 9	700	780	320 717	3 568 432	7 103 072	10 741 552

На рис. 2 представлено время обучения рассматриваемых классификаторов. Поскольку разброс значений велик, визуализация была выполнена в виде двух графиков. На рис. 2, а показано время обучения решающего дерева, наивного байесовского классификатора и логистической регрессии. Для большей наглядности этот график выполнен в логарифмическом масштабе. Видно, что решающее дерево обучается дольше всех из этой тройки. Но, как показывает рис. 2, б, случайный лес обучается еще дольше, причем время обучения случайного леса тем больше, чем больше в нем решающих деревьев.

Визуализация времени классификации выполнена аналогично, однако логарифмический масштаб не использовался, поскольку разброс значений не столь велик. Как видно из рис. 3, а и 3, б, случайный лес проигрывает остальным рассматриваемым моделям машинного обучения и по времени классификации.

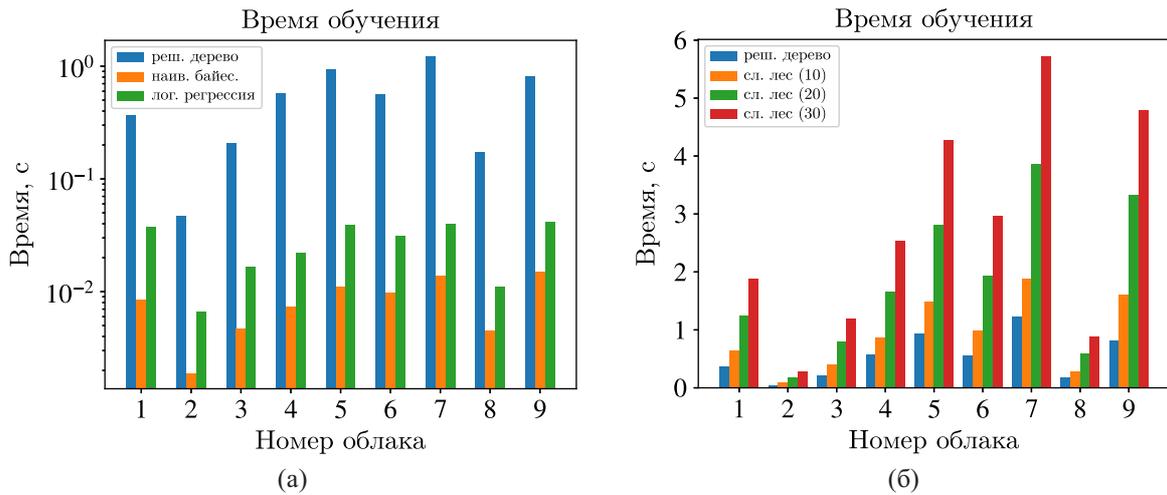


Рис. 2. Время обучения классификаторов: а) логистическая регрессия, наивный классификатор Байеса, решающее дерево; б) решающее дерево, случайный лес с 10, 20, 30 деревьями

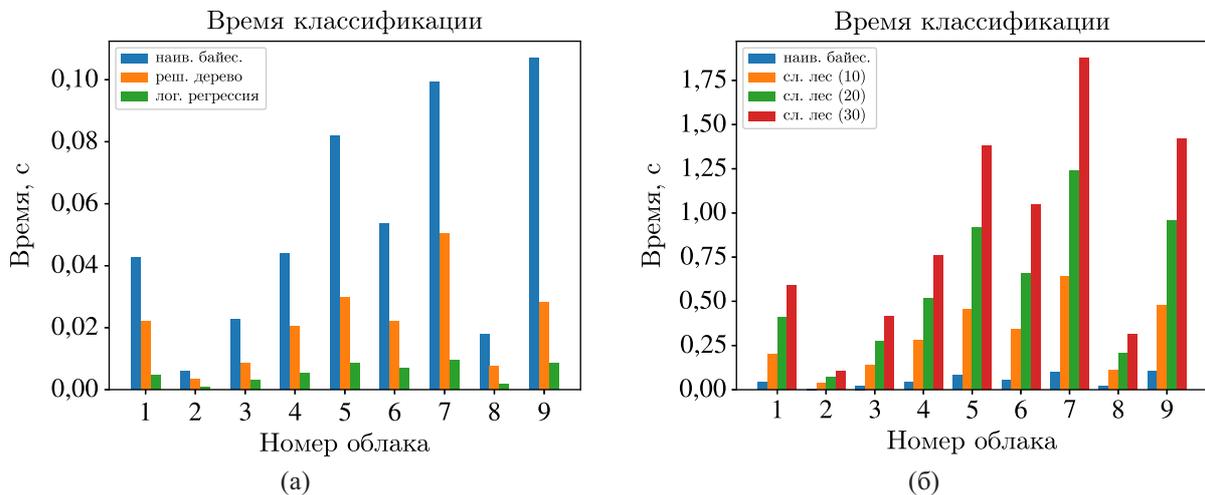


Рис. 3. Время классификации: а) с помощью логистической регрессии, наивного классификатора Байеса, решающего дерева; б) с помощью решающего дерева, случайного леса с 10, 20, 30 деревьями

На рис. 4, а представлена сбалансированная точность сегментации исследуемых классификаторов, за исключением случайного леса. Кроме уже рассмотренных нами решающего дерева, наивного байесовского классификатора и логистической регрессии (из которых решающее дерево имеет наилучшую сбалансированную точность), представлены результаты для уточненного решающего дерева логистической регрессии. Это двухступенчатый классификатор, в котором сбалансированная точность решающего дерева повышается за счет уточнения класса «ствол» с помощью логистической регрессии. Уточнение происходит следующим образом: выполняется бинарная классификация точек облака с помощью решающего дерева, после чего точки, отнесенные к листве, классифицируются с помощью логистической регрессии. На большинстве рассматриваемых стереоизображениях сбалансированная точность уточненного таким образом решающего дерева выше, чем у решающего дерева, не использующего уточнения результатов классификации.

Сравнение сбалансированной точности уточненного решающего дерева со случайным лесом показано на рис. 4, б. Видно, что уточненное решающее дерево выполняет классификацию

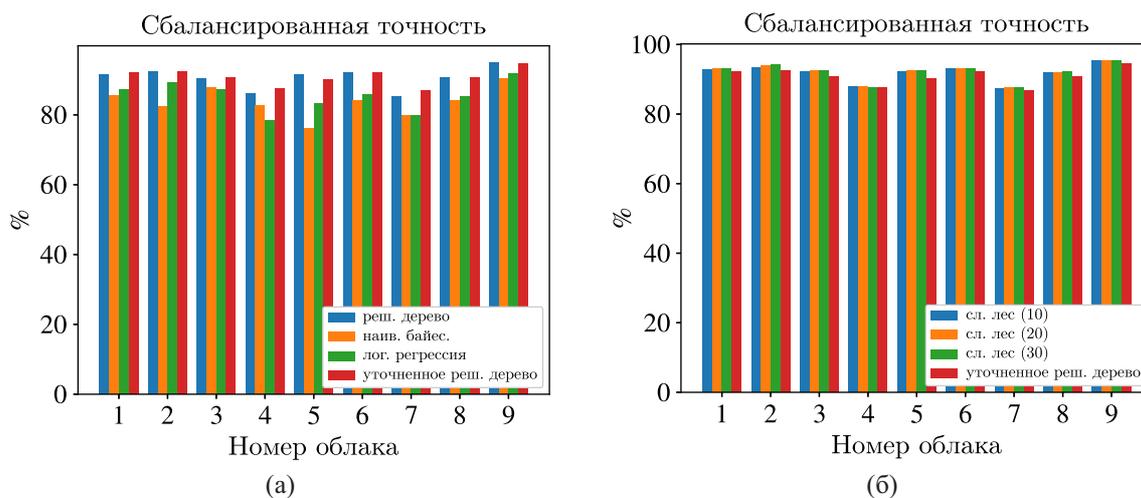


Рис. 4. Сбалансированная точность при классификации: а) с помощью логистической регрессии, уточненной логистической регрессии, наивного классификатора Байеса, решающего дерева; б) с помощью случайного леса (10, 20, 30 деревьев) и уточненной логистической регрессии

хуже, однако проигрыш совсем невелик. При этом по остальным показателям (время обучения, время классификации и потребление оперативной памяти), которые не превышают соответствующих суммарных показателей решающего дерева и логистической регрессии, двухступенчатый классификатор выигрывает у случайного леса.

Заключение

В работе был рассмотрен вопрос поиска классификатора с наименьшими требованиями к объему используемой оперативной памяти для задачи сегментации трехмерных точечных изображений лесных деревьев. Популярный в этой области случайный лес является достаточно медленным при обучении и предсказании класса и требовательным к объему памяти. Вполне естественно предположить, что такие модели машинного обучения, как логистическая регрессия и наивный байесовский классификатор, более экономичны в смысле расхода памяти. Основной вопрос заключается в том, насколько худшую точность мы получим, используя данные классификаторы вместо случайного леса.

Проведенные вычислительные эксперименты показали, что проигрыш по точности действительно имеет место, но не велик. При этом логистическая регрессия и наивный байесовский классификатор, а также решающее дерево не только более экономичны по памяти, но и быстрее при обучении и предсказании класса.

Особенно интересным был эксперимент с комбинацией решающего дерева и логистической регрессией, когда результат сегментации, выполненной первым классификатором, уточняется с помощью второго. Идея такого комбинирования была навеяна схожими примерами использования конвейеров алгоритмов обработки облаков точек, известными из литературы. В итоге классификатор, сочетающий решающее дерево и логистическую регрессию, на большинстве тестовых изображений показал наилучшую точность из всех экономичных классификаторов, лишь немного отстав от случайного леса и значительно обойдя случайный лес по времени обучения и предсказания при заметно более экономичном потреблении ОЗУ.

В дальнейшем планируется провести более полное исследование комбинированных классификаторов в задаче сегментации облаков точек.

Список литературы (References)

- Alfio V.S., Pepe M., Costantino D.* The use of random forest for the classification of point cloud in urban scene // *International Journal of Engineering Trends and Technology*. — 2024. — Vol. 72, No. 3. — P. 1–9.
- Alshawabkeh Y., El-Khalili M., Almasri E., Bala'awi F., Al-Massarweh A.* Heritage documentation using laser scanner and photogrammetry. The case study of Qasr Al-Abidit, Jordan // *Digital Applications in Archaeology and Cultural Heritage*. — 2020. — Vol. 16. — e00133.
- Domingos P., Pazzani M.* On the optimality of the simple Bayesian classifier under zero-one loss // *Machine Learning*. — 1997. — Vol. 29. — P. 103–130.
- Hu C., Kong L., Lv F.* Application of 3D laser scanning technology in engineering field // *E3S Web Conf.*, IAECST 2020. — 2021. — Vol. 233. — 04014.
- Hyypä J., Holopainen M., Olsson H.* Laser scanning in forests // *Remote Sensing*. — 2012. — Vol. 4, No. 10. — P. 2919–2922.
- Liang X., Kankare V., Hyypä J., Wang Y., Kukko A., Haggrén H., Yu X., Kaartinen H., Jaakkola A., Guan F., Holopainen M., Vastaranta M.* Terrestrial laser scanning in forest inventories // *ISPRS Journal of Photogrammetry and Remote Sensing*. — 2016. — Vol. 115. — P. 63–77.
- Lin W., Fan W., Liu H., Xu Y., Wu J.* Classification of handheld laser scanning tree point cloud based on different KNN algorithms and random forest algorithm // *Forests*. — 2021. — Vol. 12. — P. 292.
- Liu Q., Ma W., Zhang J., Liu Y., Xu D., Wang J.* Point-cloud segmentation of individual trees in complex natural forest scenes based on a trunk-growth method // *J. For. Res.* — 2021. — Vol. 32. — P. 2403–2414.
- Park J.K., Lee K.W.* Application of 3D laser scanning data for forest survey and timber volume estimation // *Sens. Mater.* — 2023. — Vol. 35, No. 9. — P. 3269–3278.
- Qi C.R., Su H., Mo K., Guibas L.J.* Pointnet: Deep learning on point sets for 3d classification and segmentation // *arXiv preprint*. — 2016. — arXiv:1612.00593
- Qi C.R., Yi L., Su H., Guibas L.J.* Pointnet++: Deep hierarchical feature learning on point sets in a metric space // *Advances in Neural Information Processing Systems (NeurIPS)*. — 2017.
- Qiu T., Du R., Spine N., Cheng L., Jiang Y.* Joint 3D point cloud segmentation using real-sim loop: from panels to trees and branches // *arXiv preprint*. — 2025. — arXiv:2503.05630v1
- Sarker S., Sarker P., Stone G., Gorman R., Tavakkoli A., Bebis G., Sattarvand J.* A comprehensive overview of deep learning techniques for 3D point cloud classification and semantic segmentation // *Machine Vision and Applications*. — 2024. — Vol. 35, No. 4.
- Sun J., Wang P., Gao Z., Liu Z., Li Y., Gan X., Liu Z.* Wood-leaf classification of tree point cloud based on intensity and geometric information // *Remote Sens.* — 2021. — Vol. 13, No. 20.
- Weiser H., Ulrich V., Winiwarter L., Esmoris A.M., Höfle B.* Manually labeled terrestrial laser scanning point clouds of individual trees for leaf-wood separation // 2024.
- Wozniak P., Krzeszowski T., Kwolek B.* Multi-domain indoor dataset for visual place recognition and anomaly detection by mobile robots // *Sci. Data*. — 2025. — Vol. 12. — P. 817.
- Yun T., An F., Li W., Sun Y., Cao L., Xue L.* A novel approach for retrieving tree leaf area from ground-based LiDAR // *Remote Sens.* — 2016. — Vol. 8. — P. 942.