

# СОДЕРЖАНИЕ

<b>ВВЕДЕНИЕ</b>	<b>4</b>
<b>1 Аналитический раздел</b>	<b>5</b>
1.1 Обзор предметной области . . . . .	5
1.1.1 Ансамблевые алгоритмы . . . . .	5
1.1.2 Обработка изображения . . . . .	7
1.2 Бэггинг . . . . .	7
1.3 Бустинг . . . . .	9
1.4 Стекинг . . . . .	11
1.5 Сравнение методов . . . . .	12

# ВВЕДЕНИЕ

Машинное обучение это отрасль искусственного интеллекта, характерной чертой которого является постоянное обучение машины и расширение аналитической модели за счет применения решений множества сходных задач.

Область применения машинного обучения расширяется с каждым днем — данный метод применяется в медицине при прогнозировании вероятности возникновения заболевания на основе известных факторов, в фото и видео технике при распознавании человека на фото и улучшении качества изображения.

При решении сложных задач часто возникает ситуация, при которой ни один из существующих алгоритмов не гарантирует требуемого качества прогнозирования. Идея ансамблевого обучения заключается в объединении отдельных алгоритмов в единую композицию, в которой сильные стороны алгоритмов будут объединены, а ошибки — взаимно компенсированы.

Целью данной работы является представление обзора и сравнения ансамблевых алгоритмов машинного обучения на основе ограниченного числа слабых экспертов для обработки фотографий, полученных со спутника.

Для достижения поставленной цели необходимо выполнить следующие задачи:

- 1) описать ансамблевые алгоритмы машинного обучения;
- 2) выбрать критерии классификации;
- 3) сравнить эти методы в контексте обработки фотографий с околоземной орбиты.

# 1 Аналитический раздел

В данном разделе проводится обзор предметной области ансамблевых алгоритмов машинного обучения и их сравнительный анализ.

## 1.1 Обзор предметной области

Далее представлен краткий анализ предметной области — определены понятия ансамблевого алгоритма и обработки изображений, рассмотрены применяемые методы.

### 1.1.1 Ансамблевые алгоритмы

Слабым учеником [?] называют алгоритм, который справляется с поставленной задачей плохо, но лучше чем случайное угадывание.

В качестве примера можно рассматривать алгоритм определения пола человека на основе его роста. Средний рост женщин в России 165 см, мужчин — 178 см. Можно сделать предположение, что все люди выше 170 см — мужчины, а ниже — женщины. Результат подобной классификации приведет к большому количеству статистических ошибок, но точность предсказания будет выше 50%.

Ансамблевые алгоритмы применяются для объединения набора слабых учеников в одну более сложную аналитическую модель — сильного ученика [?], обеспечивающего требуемый уровень точности выполнения задачи.

Наиболее известны и широко применимы такие ансамблевые алгоритмы, как: *бэггинг*, *бустинг*, *стекинг*.

Для оценки качества работы алгоритма  $a$  можно использовать [?] следующую формулу ошибки  $Q(a)$ :

$$Q(a) = \mathbb{E}_x \mathbb{E}_{X, \epsilon} [y(x, \epsilon) - a(x, X)]^2, \quad (1)$$

где

- 1)  $X$  — обучающая выборка;
- 2)  $x$  — точка из тестового множества;
- 3)  $a(x, X)$  — значение в точке  $x$  алгоритма, обученного на выборке  $X$ ;
- 4)  $y = f(x) + \epsilon$  — целевая зависимость, которую можно измерить с точностью до случайно шума  $\epsilon$ ;
- 5)  $\mathbb{E}_x$  — математическое ожидание  $x$ ;
- 6)  $\mathbb{E}_{X, \epsilon}$  — среднее по всем обучающим выборкам  $X$  и случайному шуму  $\epsilon$ .

Часто используемым инструментом для анализа работы ансамблевых алгоритмов является разложение *bias-variance decomposition*, определяющее представление  $Q(a)$  в виде трех компонент:

$$Q(a) = \mathbb{E}_x \text{bias}_X^2 a(x, X) + \mathbb{E}_x \mathbb{V}_X[a(x, X)] + \sigma^2, \quad (2)$$

где

- 1)  $\text{bias}_X a(x, X) = f(x) - \mathbb{E}_X[a(x, X)]$  — смещение предсказания алгоритма в точке  $x$  относительно истинной зависимости  $f$ ;
- 2)  $\mathbb{V}_X[a(x, X)] = \mathbb{E}_X [a(x, X) - \mathbb{E}_X[a(x, X)]]^2$  — разброс предсказаний алгоритма в зависимости от обучающей выборки  $X$ ;
- 3)  $\sigma^2 = \mathbb{E}_x \mathbb{E}_\epsilon [y(x, \epsilon) - f(x)]^2$  — неустранимый шум в данных.

### 1.1.2 Обработка изображения

Изображение [?] — это информация, пригодная для визуального восприятия.

Обработка изображения [?] — любые манипуляции с графической информацией для достижения определенных целей. По типам цели, с которой производятся манипуляции, выделяют следующие типы обработки изображений:

- 1) визуальное улучшение качества изображения (удаление шумов);
- 2) классификация — распознавание образов и объектов по фотографии;
- 3) проведение измерений на изображении (анализ интерферограмм, рентген снимков и т.д.).

В контексте данной работы возможность обработки фотографий с целью удаления шумов и решения задач классификации будут являться критериями сравнения ансамблевых алгоритмов.

Мультиспектральная визуализация (Multispectral Imaging, MSI) [?] — метод обработки фотографий, подразумевающий захват изображений в нескольких дискретных диапазонах длин волн. Данная технология позволяет производить более качественную обработку фотографий, используя для этого информацию, находящуюся за пределами цветового восприятия человеческого глаза.

Частое применение технологии MSI — улучшение качества и анализ фотографий, полученных с нескольких камер спутника, снимающих в разным диапазонах цветового спектра. Возможность использования мультиспектральной визуализации также будет являться критерием сравнения ансамблевых алгоритмов.

## 1.2 Бэггинг

Бэггинг [?] — ансамблевый алгоритм, подразумевающий обучение ограниченного числа моделей на разных подвыборках выборки  $X$  и усреднение их результата.

Пусть выборка  $X$  состоит из  $n$  элементов. Случайным образом выбирается (с возвращением)  $n$  элементов из выборки. Таким образом получается подвыборка  $X_1$ , которая содержит не все элементы  $X$ , а некоторые содержит больше одного раза. На подвыборке  $X_1$ , используя некоторый алгоритм, обучается модель  $b_1$ . Аналогичным образом создается подвыборка  $X_i$ , на которой, используя тот же алгоритм, обучается модель  $b_i$ , где  $i \in [1, k]$ ,  $k$  — количество обучаемых моделей, которые называют базовыми алгоритмами. На финальном этапе предсказания моделей агрегируются в общий ответ. При решении задачи регрессии, предсказания моделей усредняются:

$$a(x) = \frac{1}{k}(b_1(x) + \dots + b_k(x)). \quad (3)$$

Схема алгоритма бэггинга представлена на рисунке 1.

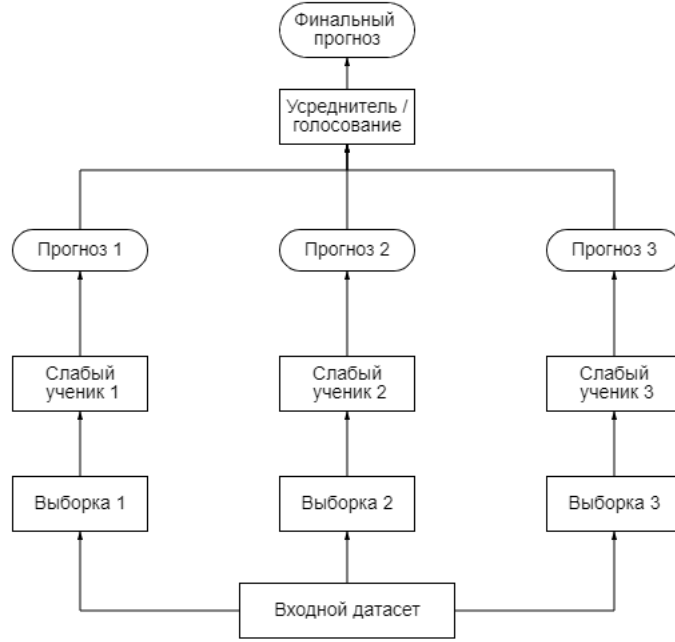


Рисунок 1 – Схема алгоритма бэггинга

При помощи преобразований определим смещение и разброс полученного ансамбля:

$$\text{bias}_X a(x, X) = f(x) - \mathbb{E}_X[a(x, X)] = \dots = \text{bias}_X b(x, X), \quad (4)$$

$$\mathbb{V}_X[a(x, X)] = \frac{1}{k^2} \sum_{i=1}^k \mathbb{V}_X b(x, X_i) + \frac{1}{k^2} \sum_{k_1 \neq k_2} \text{cov}(b(x, X_{k_1}), b(x, X_{k_2})). \quad (5)$$

Если предположить, что базовые алгоритмы некоррелированы, выра-

жение (5) приобретает следующий вид:

$$\mathbb{V}_X[a(x, X)] = \frac{1}{k} \mathbb{V}_X b(x, X). \quad (6)$$

Таким образом была получена модель, смещение которой равно смещению базового алгоритма. Разброс такой модели в  $k$  раз меньше разброса базового алгоритма.

Алгоритм бэггинга чаще всего применяется для задач регрессии и классификации, когда выходные параметры базовых алгоритмов являются, как правило, числовыми и поддаются несложным математическим методам (суммирование, усреднение и т.д.).

Таким образом, в контексте поставленной задачи, данный ансамблевый алгоритм подходит, например, для задач классификации наземной техники или распознавания зданий по фотографии, однако задача улучшения качества изображения и удаления шумов не позволяет такой же простой агрегации ответов отдельных моделей. Мультиспектральная обработка в случае использования бэггинга также недоступна или требует сложной реализации. Примером ансамблевого алгоритма, подразумевающего более сложную агрегацию ответов базовых алгоритмов является стекинг, рассмотренный ниже).

### 1.3 Бустинг

Бустинг [?] — ансамблевый метод, который, в отличие от бэггинга, состоит в последовательной адаптации нескольких слабых учеников: базовые алгоритмы выполняют предсказание последовательно и каждый следующий обучается так, чтобы уменьшить погрешность, полученную в результате работы предыдущих алгоритмов. Пример схемы обработки фотографии для избавления от шумов с использованием бустинга [?] представлен на рисунке 2.

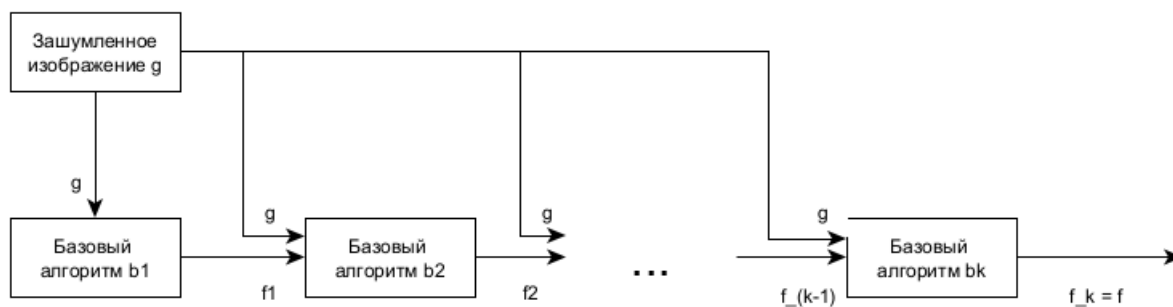


Рисунок 2 – Пример схемы обработки фотографии с использованием бустинга.  $g$  — зашумленное изображение,  $b_i$  —  $i$ -ый базовый алгоритм,  $f_i$  — результат работы  $i$ -го базового алгоритма,  $f$  — результат работы ансамбля, очищенное изображение.

В ансамбле, представленном на рисунке 2 первый базовый алгоритм получает на вход зашумленное изображение  $g$ , каждый следующий алгоритм — результат работы предыдущего алгоритма и  $g$ . Результат работы последнего базового алгоритма является результатом работы всего ансамбля.

При такой организации ансамбля каждый следующий базовый алгоритм направлен на улучшение выходного результата путем уменьшения остаточного шума и восстановления из оригинального изображения информации, потерянной при обработке предыдущим алгоритмом. Таким образом, оригинальное зашумленное изображение подается на вход каждому базовому алгоритму для избежания потери информации в ходе удаления шума.

Данный ансамблевый алгоритм позволяет получить достойное качество изображения за счет объединения сильных сторон слабых учеников и снижения погрешностей их работы путем восстановления информации, потерянной в ходе работы предыдущей модели.

Минусом данного подхода является невозможность параллельного обучения базовых алгоритмов, так как результат работы каждой модели зависит от результатов работы всех предыдущих моделей. Мультиспектральная обработка становится сложно реализуемой или невозможной.



## 1.4 Стекинг

Стекинг [?] — еще один ансамблевый алгоритм машинного обучения. В отличие от бэггинга и бустинга, подходящими только для объединения однородных (принадлежащих одному семейству алгоритмов) моделей, стекинг позволяет объединять разнородные алгоритмы.

Разнородные базовые алгоритмы обучаются на одной и той же обучающей выборке. Для объединения результатов базовых алгоритмов в единый ответ используется мета-модель (мета-ученик), обучаемая на результатах работы базовых алгоритмов.

Схема алгоритма стекинга представлена на рисунке 3.

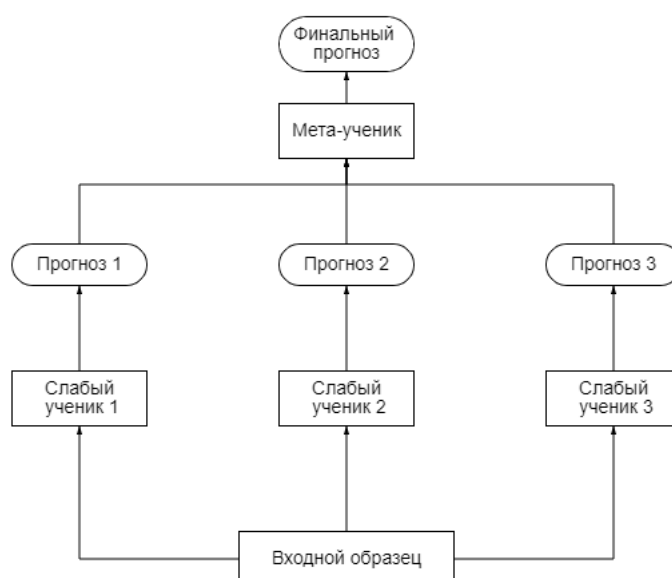


Рисунок 3 – Схема алгоритма стекинга

Главным преимуществом стекинга является возможность применения неоднородных базовых алгоритмов. При использовании мультиспектральной визуализации различные диапазоны длин волн могут быть обработаны разнородными моделями. Общий ответ создается на основе алгоритма сложения слоев (layer stacking) [?] и выполняется обученной мета-моделью.

## 1.5 Сравнение методов

Сравним рассмотренные выше ансамблевые алгоритмы, разобрав их по следующим критериям:

- 1) параллельность — возможность обучать базовые алгоритмы параллельно;
- 2) классификация — применимость в задачах классификации объектов по фотографии;
- 3) улучшение качества — применимость в задачах удаления шумов с изображения;
- 4) MSI — применимость мультиспектральной обработки изображения.

Разбор соответствующих критериев представлен в таблице 1.

Таблица 1 – Сравнительная таблица ансамблевых алгоритмов

Алгоритм	Параллельность	Классификация	Улучшение качества	MSI
Бэггинг	+	+	-	-
Бустинг	-	+	+	-
Стекинг	+	+	+	+

По результатам сравнения с использованием выбранных критериев, алгоритм стекинга является наиболее универсальным для реализации ансамблей, выполняющих различные виды обработки изображений, однако нельзя назвать данный алгоритм однозначно лучшим.

Согласно данным, приведенным в работе [?], бустинг показывает наилучший результат при условии, что в данных немного шума, однако требует значительно больше времени на обучение базовых алгоритмов.

Помимо показателей точности предсказания также следует учитывать следующие критерии выбора ансамблевого алгоритма.

- 1) Применимость к конкретной задаче: рассмотренные методы различаются в возможностях в зависимости от характера обрабатываемых данных (высокая зашумленность, высокая размерность, многоуровневая классификация и т.д.). Рекомендуется предварительная оценка решаемой задачи для выбора соответствующего ансамблевого алгоритма.
- 2) Вычислительная стоимость: важными критериями выбора алгоритма могут являться вычислительная стоимость и требуемое время предсказания нового запроса. Данные критерии особенно актуальны для систем реального времени.
- 3) Программная доступность: различные платформы предоставляют собственные реализации алгоритмов машинного обучения. Возможность применения одного и того же алгоритма в нескольких приложениях может быть важным параметром при выборе ансамблевого алгоритма.
- 4) Легкость в применении: пользователи предпочитают понимать, какими методами можно настроить используемые модели, и, следовательно, чаще отдают предпочтение моделям, обладающим четким набором контролируемых параметров.

Таким образом, невозможно назвать однозначно лучший алгоритм ансамблевого обучения. Выбор наиболее оптимального алгоритма может быть произведен только на основе анализа конкретной задачи.