Отношение сигнал/шум в астрофизических наблюдениях

Лекция 11

Методы накопления сигнала и объединения данных. Stacking, усреднение, биннинг и их влияние на S/N

КазНУ имени аль-Фараби

Цель и план лекции

Цель

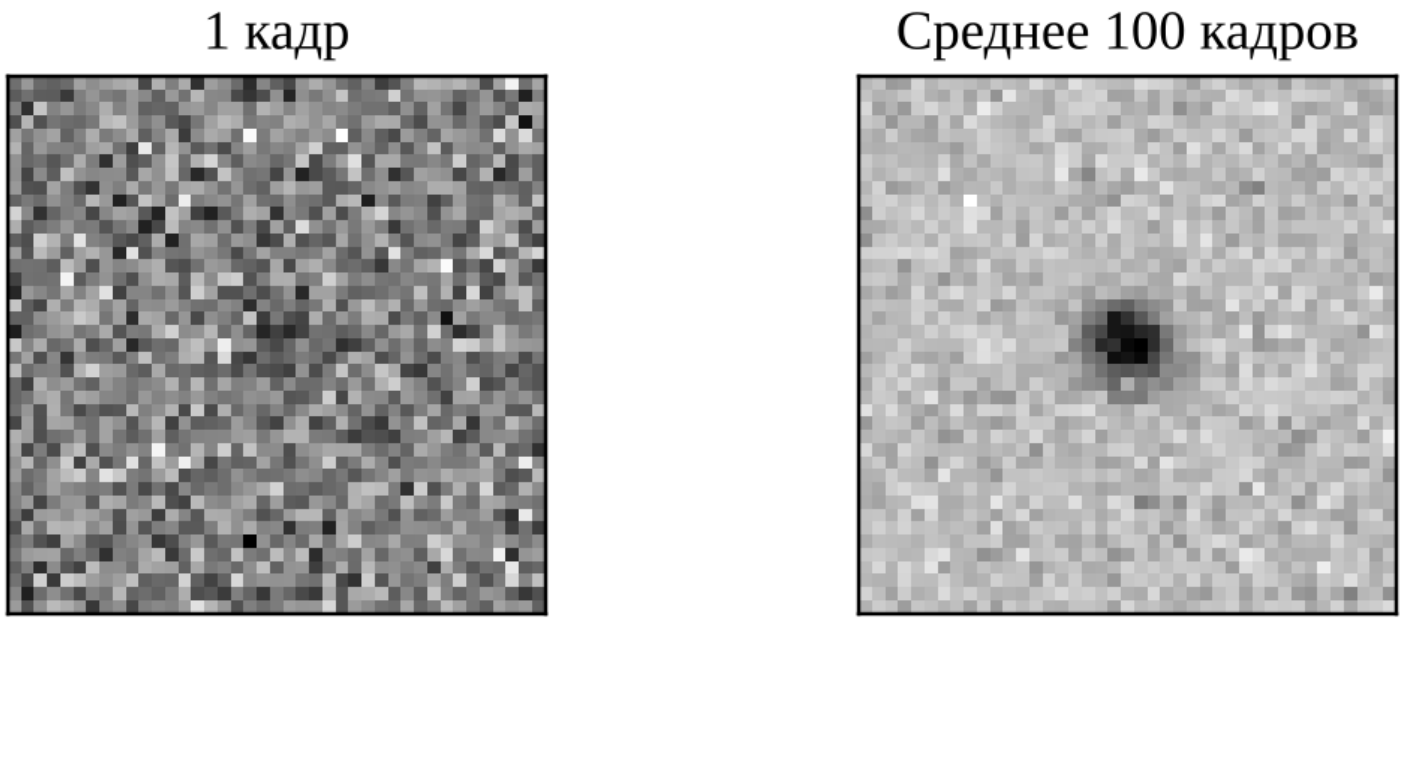
Изучение методов объединения наблюдений и расчёта S/N, достигаемого при их применении, а также умение выбирать способ объединения с учётом характера шума и наличия выбросов.

1. Введение
2. Основные методы объединения
3. Биннинг
4. Выравнивание и ресэмплинг
5. Ограничения накопления
6. Объединение данных разных приборов и эпох

1 Введение

- Объединение нескольких измерений является одним из основных способов повышения качества данных.

1 Введение (продолжение)



Изображение источника в одиночном кадре, среднее 100 кадров и зависимость прироста S/N от числа кадров

2 Основные методы объединения

Среднее

Для N независимых кадров одинакового качества шум уменьшается в √N раз. Этот метод наиболее эффективен для гауссова шума, но чувствителен к выбросам.

Медиана

Обеспечивает защиту от выбросов, но при нормальном шуме дисперсия больше в π/2 раз, поэтому S/N меньше примерно в 1,25 раза.

Среднее с отсечением

Итеративно исключаются значения, отклоняющиеся более чем на kσ от текущей оценки (sigma clipping).

Минимум–максимум

Отбрасываются наибольшие и наименьшие значения. Метод прост, но не использует сведения о шуме.

Взвешенное среднее

Весами служат обратные дисперсии, чтобы кадры лучшего качества вносили больший вклад.

2 Основные методы объединения (продолжение)

- При оптимальных весах:

$$w_i \propto 1 / \sigma_i^2, \quad S/N^2 = \Sigma_i (S/N)_i^2.$$

- Таким образом, квадраты S/N независимых наблюдений складываются. Для изображений необходимо учитывать также различие качества: кадры с лучшим seeing содержат больше пиксельной информации для точечных объектов.

$$w_i \propto F_i^2 / (\sigma_i^2 \cdot n_{eff,i}).$$

3 Биннинг

Биннинг объединяет соседние пиксели или значения.

Аппаратный биннинг

В ПЗС заряд суммируется до считывания, что вносит шум считывания однократно для группы пикселей.

Программный биннинг

Значения складываются после считывания, поэтому шум считывания суммируется в квадратуре.

Биннинг спектров и временных рядов

Несколько соседних отсчётов объединяются с уменьшением разрешения. Применяется, если значение S/N на элемент низкое, а разрешение избыточно.

Необходимо учитывать, что биннинг снижает разрешение и может ухудшить результат при неоднородности данных.

4 Выравнивание и ресэмплинг

- Перед объединением изображений их необходимо совместить в единой системе координат.
- Закай и Офек (2017) предложили оптимальный способ объединения изображений с различной PSF, который сводится к взвешиванию в пространстве Фурье и обеспечивает оптимальность как для обнаружения источников, так и для измерений.

5 Ограничения накопления

- Накопление эффективно, пока доминирует случайный шум. Систематические эффекты ограничивают улучшение: ошибка плоского поля, предельная спутанность, неточность выравнивания приводят к тому, что зависимость $\sigma \propto N^{-1/2}$ перестаёт выполняться.
- Отдельным методом является статистическое накопление по объектам (stacking): кадры, центрированные на положениях множества слабых источников из каталога, суммируются, чтобы определить их средние свойства, даже если каждый объект в отдельности не обнаружен.

5 Сравнение методов объединения на нормальном шуме

Листинг 11.1.

```
import numpy as np

rng = np.random.default_rng(11)
N, trials, sig = 25, 20000, 1.0
X = rng.normal(0, sig, (trials, N))

print(f"sigma/sqrt(N) = {sig / np.sqrt(N):.4f}")
print(f"Среднее:  {X.mean(1).std():.4f}")
print(f"Медиана:  {np.median(X, 1).std():.4f}")
print(f"Теория медианы: {1.2533 * sig / np.sqrt(N):.4f}")

s = np.array([1, 1, 1, 2, 2, 3, 4, 4, 5, 5] * 2)[:20].astype(float)
Y = rng.normal(0, s, (trials, s.size))
w = 1 / s**2
print(f"Неоднородные кадры: простое среднее "
      f"{Y.mean(1).std():.4f}, взвешенное "
      f"{((Y * w).sum(1) / w.sum()).std():.4f}, теория "
      f"{w.sum()**-0.5:.4f}")
```

5 Результат: Сравнение методов объединения на нормальном шуме

```
sigma/sqrt(N) = 0.2000
Среднее:  0.1983
Медиана:  0.2480
Теория медианы: 0.2507
Неоднородные кадры: простое среднее 0.7187, взвешенное 0.3631, теория 0.3620
```

5 Устойчивость при наличии космических лучей

Листинг 11.2.

```
def clipped_mean(a, k=3.0, it=5):
    a = a.copy()
    m = np.ma.masked_invalid(a)
    for _ in range(it):
        c = np.ma.median(m, axis=1)[:, None]
        s = np.ma.std(m, axis=1)[:, None]
        m = np.ma.masked_where(np.abs(m - c) > k * s, m)
    return m.mean(axis=1).filled(np.nan)

N, trials = 25, 4000
X = rng.normal(100.0, 5.0, (trials, N))
hits = rng.random((trials, N)) < 0.02
X[hits] += rng.exponential(500, hits.sum())

def rep(name, a):
    print(f"{name:18} смещение {a.mean() - 100:7.3f}  СКО {a.std():.3f}")

rep("Среднее", X.mean(1))
rep("Медиана", np.median(X, 1))
rep("С отсечением 3σ", clipped_mean(X))
```

5 Результат: Устойчивость при наличии космических лучей

```
Среднее            смещение   9.595  СКО 19.057
Медиана            смещение   0.114  СКО 1.260
С отсечением 3σ    смещение   0.007  СКО 1.024
```

Среднее значение смещено и имеет высокий разброс из-за космических лучей. Метод с отсечением сохраняет низкий разброс и почти не смещён, в этом случае превосходит медиану.

6 Объединение данных разных приборов и эпох

- На практике приходится объединять наблюдения, полученные при разной глубине, разном разрешении и даже разными приборами.
- Для переменных объектов простое суммирование недопустимо: изменения блеска размывают результат, а положение быстрых объектов (астероидов, искусственных спутников) меняется от кадра к кадру.
- При накоплении спектров следует учитывать различия лучевых скоростей: для наблюдений с разными датами спектры необходимо перевести в барицентрическую систему.
- Эффективность обработки зависит от правильного учёта масок дефектов: каждому пикселю итогового кадра необходимо присваивать своё значение эффективной экспозиции.
- Упражнение: смоделируйте 20 кадров с различным качеством и сравните итоговые значения S/N при равных и оптимальных весах, а также при отборе лучших кадров.

7 Разбор примера

1. Пусть имеются 3 кадра с S/N = 10, 20 и 30.

2. При оптимальном взвешивании итоговое значение $S/N = \sqrt{(100 + 400 + 900)} = \sqrt{1400} \approx 37{,}4$.

3. Если выполнять простое усреднение с равными весами и предположить, что сигнал равен единице, а шумы равны 0,1, 0,05 и 0,033, то шум среднего равен $\sqrt{(0{,}01 + 0{,}0025 + 0{,}0011)}/3 \approx 0{,}0389$, что даёт $S/N \approx 25{,}7$.

4. Таким образом, простое усреднение в данном случае оказывается хуже даже лучшего отдельного кадра с S/N = 30.

5. Биннинг 2×2 в ПЗС: сигнал увеличивается в 4 раза.

6. При режиме шума считывания дисперсия остаётся равной RN^2, поэтому S/N увеличивается в 4 раза.

7. При фоновом режиме дисперсия увеличивается в 4 раза, а S/N – только в 2 раза.

8 Вопросы для обсуждения на занятии

Выбор метода

Какой способ объединения следует применять при очень малом и большом количестве кадров?

Веса

Какие величины следует использовать для весов кадров с разными условиями?

Ограничения

Как установить, что накопление больше не улучшает результат?

Контрольные вопросы

1. Почему накопление данных увеличивает S/N?
2. Сравните эффективность среднего и медианы.
3. Что такое среднее с отсечением? Какие параметры оно имеет?
4. Какие веса являются оптимальными при объединении кадров разного качества?
5. Чем аппаратный биннинг отличается от программного?
6. Как ресэмплинг влияет на статистику шума?
7. Какие эффекты ограничивают накопление сигнала?
8. Что такое статистическое накопление по каталогу объектов?

Список литературы

1. Fruchter A.S., Hook R.N. Drizzle: A method for the linear reconstruction of undersampled images // Publications of the Astronomical Society of the Pacific. – 2002. – Vol. 114. – P. 144–152.
2. Zackay B., Ofek E.O. How to coadd images. I. Optimal source detection and photometry using ensembles of images // The Astrophysical Journal. – 2017. – Vol. 836. – Art. 187.
3. Bertin E., Arnouts S. SExtractor: Software for source extraction // Astronomy and Astrophysics Supplement Series. – 1996. – Vol. 117. – P. 393–404.
4. Howell S.B. Handbook of CCD Astronomy. – 2nd ed. – Cambridge: Cambridge University Press, 2006.
5. Ivezić Ž., Connolly A.J., VanderPlas J.T., Gray A. Statistics, Data Mining, and Machine Learning in Astronomy: A Practical Python Guide for the Analysis of Survey Data. – Updated ed. – Princeton: Princeton University Press, 2020.