АЛГОРИТМЫ И СТРУКТУРЫ ДАННЫХ

ЛЕКЦИЯ 8

Жадные алгоритмы. Задача составления расписания. Задача о рюкзаке

Для жадного правила нужна гарантия: удачный пример её не заменяет.

08

ИЗ 15 ЛЕКЦИЙ



МОДУЛЬ 3 Динамическое программирование

1 КУРС / 2 АКАДЕМИЧЕСКИХ ЧАСА / PYTHON 3

Цель и содержание лекции

ЦЕЛЬ И РЕЗУЛЬТАТЫ ОБУЧЕНИЯ

формулировать жадное правило выбора, доказывать его корректность для отбора непересекающихся занятий и строить контрпримеры для рюкзака с неделимыми предметами.

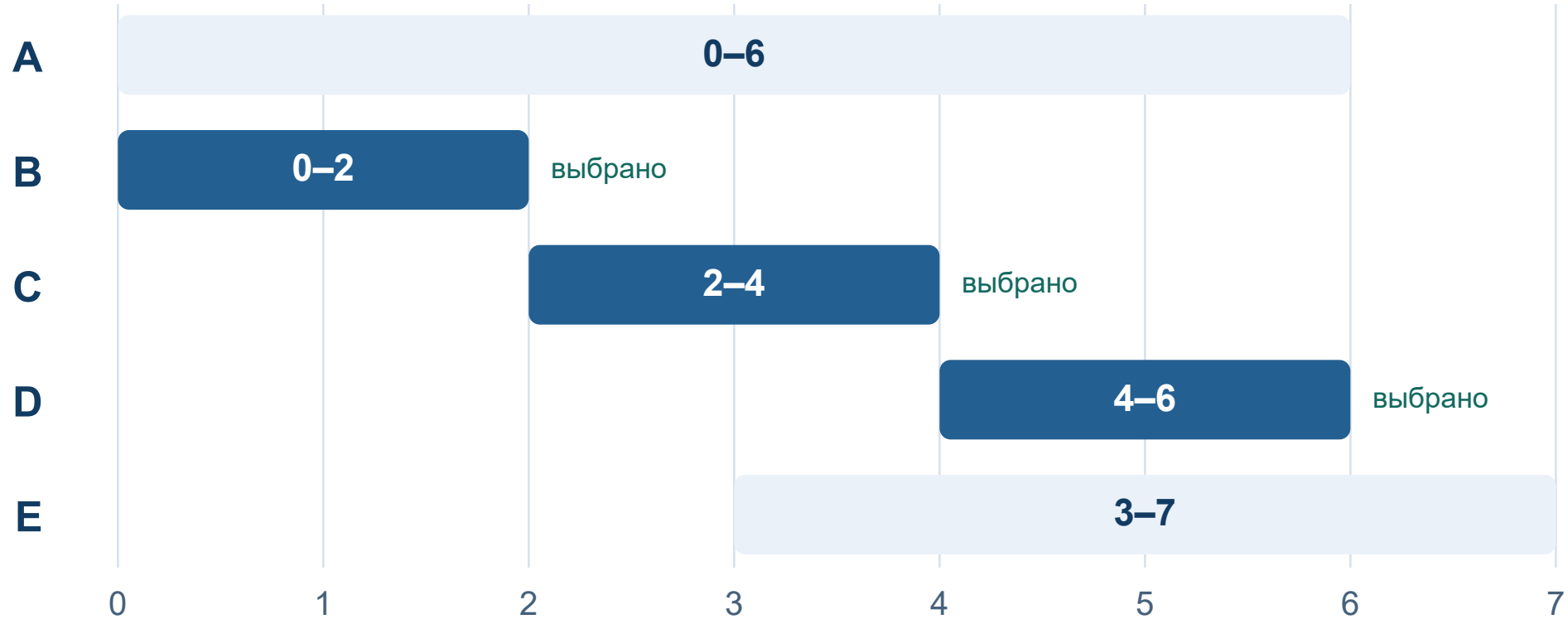
- Обосновывать выбор занятия с самым ранним окончанием.
- Различать дробный рюкзак и рюкзак 0/1.
- Находить контрпример к жадной стратегии.

ПЛАН ЛЕКЦИИ

1. **Локальный выбор в задачах оптимизации**
2. **Отбор занятий по времени окончания**
3. **Доказательство методом обмена**
4. **Рюкзак 0/1 и дробный рюкзак**
5. **Реализация на Python и применение к покрытию множества**

Раннее окончание оставляет больше времени

СХЕМА



При одинаковой ценности занятий правило раннего окончания выбирает B, C и D.

Локальный выбор в задачах оптимизации

01 ТЕМА

В задаче оптимизации требуется выбрать допустимое решение и сделать некоторую величину максимально большой или маленькой. Для расписания эта величина может быть числом занятий, а для рюкзака — суммой ценностей предметов. Ограничения определяют допустимые решения. Занятия в одной аудитории не пересекаются, а суммарная масса предметов не превышает вместимость рюкзака.

Жадный алгоритм строит решение по шагам. На каждом шаге он выбирает лучший из доступных вариантов по заданному правилу и продолжает работу с оставшимися возможностями. Такой выбор обычно окончателен. Для одной и той же задачи можно придумать несколько правил, и они могут давать разные ответы. Например, занятия можно выбирать по началу, длительности или окончанию. Чтобы утверждать, что выбранное правило всегда приводит к глобальному оптимуму, нужно это доказать. Для опровержения достаточно одного допустимого входа, на котором результат хуже оптимального.

Отбор занятий по времени окончания

02 ТЕМА

Есть одна аудитория. Начало и конец каждого занятия заранее фиксированы, занятия имеют положительную длительность и одинаково важны. Нужно провести как можно больше занятий. Урок, начинающийся в момент окончания предыдущего, допустим. Иными словами, занятия занимают полуоткрытые интервалы `[начало, конец)`.

Рассмотрим пример из главы 8. Записи уже отсортированы по времени окончания. Запоминаем конец последнего принятого занятия. Каждое следующее занятие добавляем, если оно начинается в этот момент или позже.

Отбор занятий по времени окончания

02 ПРОДОЛЖЕНИЕ

Занятие	Начало	Конец	Решение
Рисование	09:00	10:00	Выбираем первым
Английский	09:30	10:30	Пересекается с рисованием
Математика	10:00	11:00	Выбираем следующим
Информатика	10:30	11:30	Пересекается с математикой
Музыка	11:00	12:00	Выбираем последней

Результат — три занятия. Проверим другой критерий. Пусть имеются интервалы `[9, 12)`, `[10, 11)` и `[11, 12)`. Выбор по самому раннему началу захватит первый интервал и оставит одно занятие. Выбор по самому раннему окончанию позволит взять два последних. При равном времени окончания двух доступных занятий любой выбор оставляет одинаковое время для следующего урока.

Доказательство методом обмена

Обозначим через `g` занятие с самым ранним окончанием. Возьмём любое оптимальное расписание и обозначим его первое занятие через `a`. Если `a` отличается от `g`, заменим `a` на `g`. Занятие `g` заканчивается не позже занятия `a`. Все последующие занятия начинались в момент окончания `a` или позже, значит, они остаются допустимыми. Число занятий не меняется. Поэтому существует оптимальное расписание, которое начинается с жадного выбора `g`.

После выбора `g` остаётся задача того же вида, где разрешены только занятия с началом не раньше окончания `g`. Если бы для этого остатка существовало расписание длиннее остатка оптимального решения, добавление `g` дало бы ещё лучшее расписание для исходной задачи. Повторяя обмен для каждого шага, получаем доказательство оптимальности всего алгоритма.

Доказательство опирается на одинаковый вклад каждого занятия. Если ввести баллы важности, обмен может изменить сумму баллов. Например, при занятиях `[9, 10)` и `[10, 11)` ценностью по 1 баллу и занятии `[9, 11)` ценностью 5 баллов правило раннего окончания принесёт 2 балла, хотя допустимый максимум равен 5.

Доказательство методом обмена

03 ПРОДОЛЖЕНИЕ

Условия модели полезно проверять на границах. Если требуется ровно 10 минут на смену групп, занятие в 10:00 уже нельзя ставить сразу после рисования, заканчивающегося в 10:00. При одинаковом для всех занятий перерыве правило сохраняется. К каждому времени окончания добавляют 10 минут и сравнивают с этим значением начало следующего занятия. Порядок по окончанию остаётся тем же. Если время подготовки зависит от пары занятий, окончание одного занятия уже не определяет все допустимые продолжения. Прежнее доказательство обменом тогда нельзя перенести без проверки.

Рюкзак 0/1 и дробный рюкзак

04 ТЕМА

В задаче о рюкзаке **0/1** каждый предмет можно взять целиком один раз или оставить. В примере из главы 8 вместимость равна 35 фунтам. Магнитофон весит 30 фунтов и стоит 3000, ноутбук весит 20 фунтов и стоит 2000, гитара весит 15 фунтов и стоит 1500. Жадное правило «сначала самый дорогой предмет» выбирает магнитофон. Остатка в 5 фунтов для следующего предмета недостаточно. Но ноутбук и гитара вместе весят ровно 35 фунтов и стоят 3500. Значит, правило может пропустить оптимальный набор.

Можно попробовать выбирать по отношению ценности к весу. Для неделимых предметов такое правило тоже иногда ошибается. При вместимости 10 даны предметы с парами «вес, ценность» `(6, 9)`, `(5, 7)` и `(5, 7)`. Самое высокое отношение у первого предмета, после его выбора остаётся только 4 единицы вместимости. Два других предмета помещаются вместе и дают 14 вместо 9.

Рюкзак 0/1 и дробный рюкзак

04 ПРОДОЛЖЕНИЕ

В **дробном рюкзаке** разрешены части предметов, а их ценность пропорциональна взятой доле. При положительных весах и неотрицательных ценностях сортировка по убыванию отношения ценности к весу даёт оптимум. Если в допустимом решении занято место материалом меньшей удельной ценности, а материал большей удельной ценности остался снаружи, можно заменить равную массу и повысить итоговую ценность. Заполняя вместимость в таком порядке, при необходимости берём долю последнего предмета. Пример с вместимостью 10 даёт 9 за первый предмет и `4/5` от ценности 7 за часть второго. Суммарная ценность равна 14,6.

Реализация на Python и применение к покрытию множества

05 ТЕМА

В примере время задано целым числом в формате ЧЧММ в пределах одного дня. Для сравнения времени такой записи достаточно. Обе функции сначала сортируют данные. Функция рюкзака демонстрирует правило выбора по цене как эвристику.

Реализация на Python и применение к покрытию множества

05 ПРИМЕР PYTHON

08_01.py / фрагмент 1 из 2

```
def select_lessons(lessons):
    selected = []
    available_at = 0
    for name, start, finish in sorted(lessons, key=lambda lesson: lesson[2]):
        if start >= available_at:
            selected.append(name)
            available_at = finish
    return selected


def choose_by_price(items, capacity):
    selected = []
    total_value = 0
    for name, weight, value in sorted(items, key=lambda item: -item[2]):
        if weight <= capacity:
            selected.append(name)
            total_value += value
            capacity -= weight
    return selected, total_value
```

Реализация на Python и применение к покрытию множества

05 ПРИМЕР PYTHON

08_01.py / фрагмент 2 из 2

```
lessons = [
    ('Музыка', 1100, 1200), ('Рисование', 900, 1000),
    ('Английский', 930, 1030), ('Математика', 1000, 1100),
    ('Информатика', 1030, 1130),
]
items = [('магнитофон', 30, 3000), ('ноутбук', 20, 2000),
         ('гитара', 15, 1500)]

print(', '.join(select_lessons(lessons)))
print(choose_by_price(items, 35))
```

РЕЗУЛЬТАТ ЗАПУСКА

```
Рисование, Математика, Музыка
(['магнитофон'], 3000)
```

Реализация на Python и применение к покрытию множества

05 ПРОДОЛЖЕНИЕ

Сортировка занимает `O(n log n)` времени, проход после неё — `O(n)`. В этой реализации `sorted` создаёт новый список, поэтому используется `O(n)` дополнительной памяти. Для уже отсортированных входных данных проход можно выполнять сразу.

Функция отбора возвращает пустой список, если занятий нет. Функция рюкзака при нулевой вместимости и положительных весах возвращает пустой набор ценностью 0. В обоих примерах выбранные элементы образуют допустимый ответ. Для проверки оптимальности одного только соблюдения ограничений недостаточно. Для маленького рюкзака можно перечислить все подмножества и сравнить их стоимость с жадным ответом.

Реализация на Python и применение к покрытию множества

05 ПРОДОЛЖЕНИЕ

В задаче о покрытии множества из книги требуется охватить все регионы как можно меньшим числом радиостанций. Для станций одинаковой стоимости естественное жадное правило выбирает станцию, покрывающую больше всего ещё не охваченных регионов. Затем покрытые регионы удаляются из множества оставшихся. Если новые регионы покрыть уже нельзя, входные станции в совокупности не покрывают всю область. При наличии покрытия правило даёт допустимое решение, размер которого может превышать оптимум. На лабораторной работе полезно сверить жадный ответ с полным перебором для маленького набора станций.

Что нужно уметь после лекции

ИТОГИ

Правило раннего окончания даёт максимальное число совместимых занятий. Метод обмена связывает каждый локальный выбор с некоторым оптимальным расписанием. Для рюкзака важно условие делимости предметов, поскольку оно определяет, можно ли обосновать выбор по удельной ценности.

> Для жадного правила нужна гарантия: удачный пример её не заменяет.

Проверьте понимание

КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Почему сравнение `start >= available_at` допускает соседние занятия?
2. Как замена первого занятия используется в доказательстве оптимальности?
3. Почему выбор предмета с самой высокой ценой может ухудшить общий результат?
4. Почему доказательство для дробного рюкзака использует возможность взять часть предмета?
5. Что означает отсутствие станции, покрывающей хотя бы один оставшийся регион?

Самостоятельная работа

ПРАКТИКА

1. Постройте контрпример для правила выбора занятия минимальной длительности. Достаточно трёх занятий.
2. Реализуйте дробный рюкзак. Проверьте пример с предметами `(6, 9)`, `(5, 7)` и `(5, 7)` и вместимостью 10.
3. Реализуйте жадное покрытие для регионов `A, B, C, D, E, F` и станций с наборами `{A, B, C, D}`, `{A, B, E}`, `{C, D, F}`. Сравните жадный ответ с оптимальным.

Источники и документация

ЧТЕНИЕ

ОСНОВНОЕ ЧТЕНИЕ

А. Бхаргава. Грокаем алгоритмы

СПб.: Питер, 2017. Глава 8. Жадные алгоритмы. С. 182–205.

> При чтении сначала вручную пройдите один пример, затем запустите код и сравните результат.