上海交通大学先进网络实验室
Advanced Network Laboratory

贪心算法

主讲：张嘉乐
IEEE试点班 2018级
tt885.github.io

2021.6

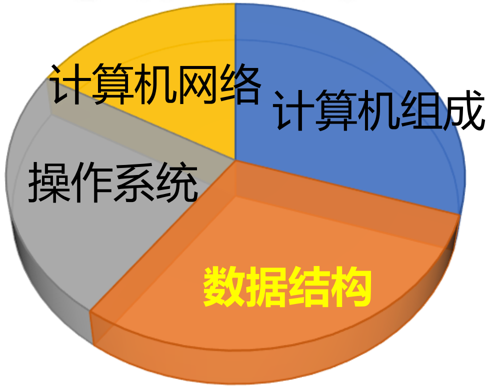
CONTENTS

算法很重要

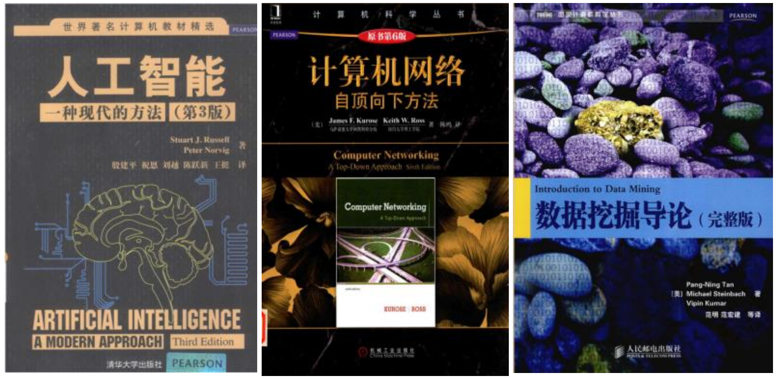
算法设计与分析是计算机科学核心研究领域

研究生面试、机考

考研专业课



其他专业课基础



互联网公司必考

Google 谷歌 Microsoft

Tencent腾讯 Alibaba 阿里巴巴

科研基础

相关竞赛

CUMCM 中国大学生数学建模竞赛

COMAP

算法是什么

欧几里得辗转相除法

Two descriptions of Euclid's algorithm

Step 1 If $n = 0$, return m and stop; otherwise go to Step 2

Step 2 Divide m by n and assign the value of the remainder to r

Step 3 Assign the value of n to m and the value of r to n. Go to Step 1.

while $n \neq 0$ do

$r \leftarrow m \bmod n$

$m \leftarrow n$

$n \leftarrow r$

return m

埃拉托色尼质数筛法

Sieve of Eratosthenes

Input: Integer $n \geq 2$

Output: List of primes less than or equal to n

for $p \leftarrow 2$ to n do $A[p] \leftarrow p$

for $p \leftarrow 2$ to $\lfloor \sqrt{n} \rfloor$ do

if $A[p] \neq 0$ //p hasn't been previously eliminated from the list

$j \leftarrow p * p$

while $j \leq n$ do

$A[j] \leftarrow 0$ //mark element as eliminated

$j \leftarrow j + p$

Example: 2 3 ~~4~~ 5 ~~6~~ 7 ~~8~~ ~~9~~ ~~10~~ 11 ~~12~~ 13 ~~14~~ ~~15~~ ~~16~~ 17 ~~18~~ 19 ~~20~~

算法解决实际问题的过程

1.提问：提出问题

2.建模：抽象出合理数学模型

3.设计：探讨算法设计技术

4.编程：给出算法形式描述

5.分析：对算法正确性和效率进行论证

6.推广：对结果进行推广

CONTENTS

贪心算法例题一览

教材风格

区间调度，区间分割

最小延迟调度

离线最优缓存

最佳找零

最小生成树

哈夫曼树

竞赛风格

策略题：分牌(NOIP2002)，过河(POJ1700)，赛马(POJ2287)

区间题：铺路(POJ2437), 排班

赶工调度：赶作业(HDU1789), 打陨石(GYM100247I)，超市(POJ1456)

计蒜客 T2158 NOIP2002

N堆纸牌，每堆a_j张，每次在相邻牌堆之间移动牌(一次可多张)，最后每堆牌数量相等

求最小移动次数，输入保证牌总数为N的倍数

输入： n ; a1,a2,……， an;

输出： 最小移动次数

样例输入： 4
9 8 17 6

样例输出： 3

样例输入： 3
9 10 11

样例输出： 2

演示：基本编程

策略 | 过河问题

POJ1700 有n人要过河，小船最多坐两人，每个人划船时间t_j，两人同乘以慢的为准求所有人过河的最小总时长

输入： n ; t1,t2,……, tn;

输出： 所有人过河的最小总时长

演示：排序

样例输入： 4
1 2 5 10

样例输出： 17

解释： 1号和2号过河(+2)，1号回来(+1)，3号和4号过河(+10),2号回来(+2),1号和2号过河(+2)。总时长17。

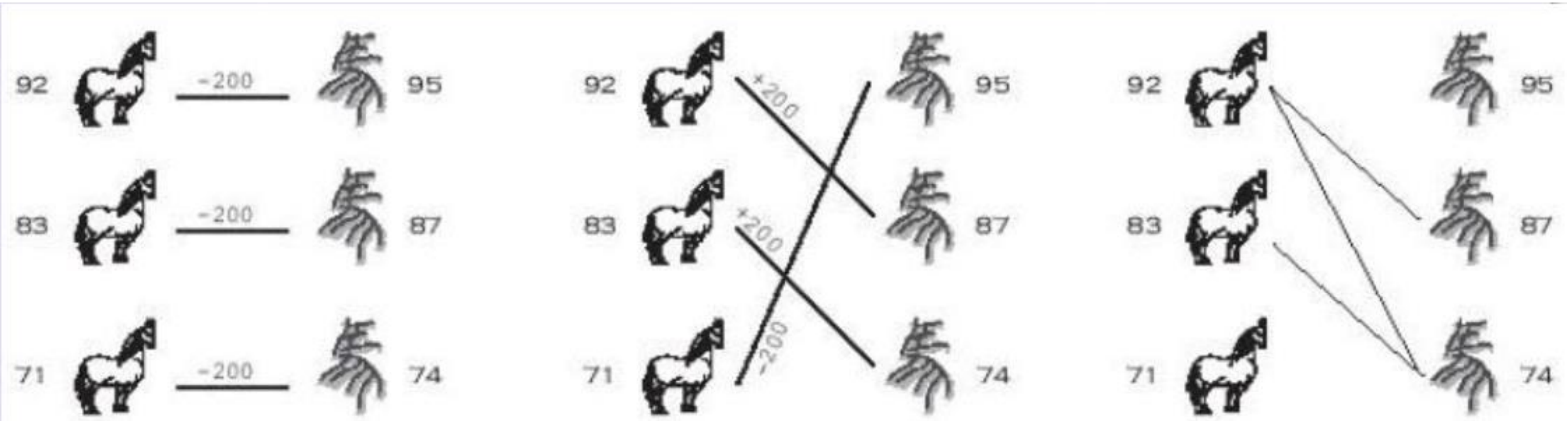
策略 ｜ 田忌赛马

《史记 - 孙子吴起列传第五》：齐使者如梁，孙膑以刑徒阴见，说齐使。齐使以为奇，窃载与之齐。齐将田忌善而客待之。忌数与齐诸公子驰逐重射。孙膑见其马足不甚相远，马有上、中、下辈。于是孙膑谓田忌曰："君第重射，臣能令君胜。"
田忌信然之，与王及诸公子逐射千金。及临质，孙子曰："今以君之下驷与彼上驷，取君上驷与彼中驷，取君中驷与彼下驷。"既驰三辈毕，而田忌一不胜而再胜，卒得王千金。

问题：诸公子不服，购良马再战。田忌和齐王各有n匹马，给出每匹马的速度值t_i，q_i
求：田忌最多净胜几场
输入：n ; t1,t2,……，tn; q1,q2,……，qn
输出：田忌最多净胜场次
样例输入 3
92 83 71
95 87 74
样例输出 1

策略 | 田忌赛马

问题： POJ2287 田忌和齐王各有n匹马，给出每匹马的速度值t_i，q_i
求田忌最多净胜几场
输入： n ; t1,t2,……，tn; q1,q2,……，qn
输出： 田忌最多净胜场次
样例输入 3
92 83 71
95 87 74
样例输出 1



牛刀： 二分图最大权匹配，最小费用最大流算法

区间 | 区间调度，区间分割

区间调度： SJOJ1360 某天有n个活动，每个时间 $[s_j, d_j]$，同一时间只能参加一个活动
求： 一个人当天能参加的最多活动数
输入： n ; s1,s2,……，sn; d1,d2,……，dn
输出： 一个人最多能参加的活动数

区间分割： 某天有n个活动，每个时间 $[s_j, d_j]$，同一时间只能参加一个活动
求： 至少多少个人才能参加完所有活动
输入： n ; s1,s2,……，sn; d1,d2,……，dn
输出： 能参加完所有活动的最少人数

POJ2437 一条直路有n个坑，每个范围 $[s_j, d_j]$，用长度L的木板铺平坑，求最少木板数.
输入： n ; s1,s2,……， sn; d1,d2,……， dn
输出： 最少的木板数

演示：排序

换壳: 考试前一天有n个同学来问问题，每个预约时间范围 $[s_j, d_j]$，有工作时长L的助教
求： 最少的助教数。

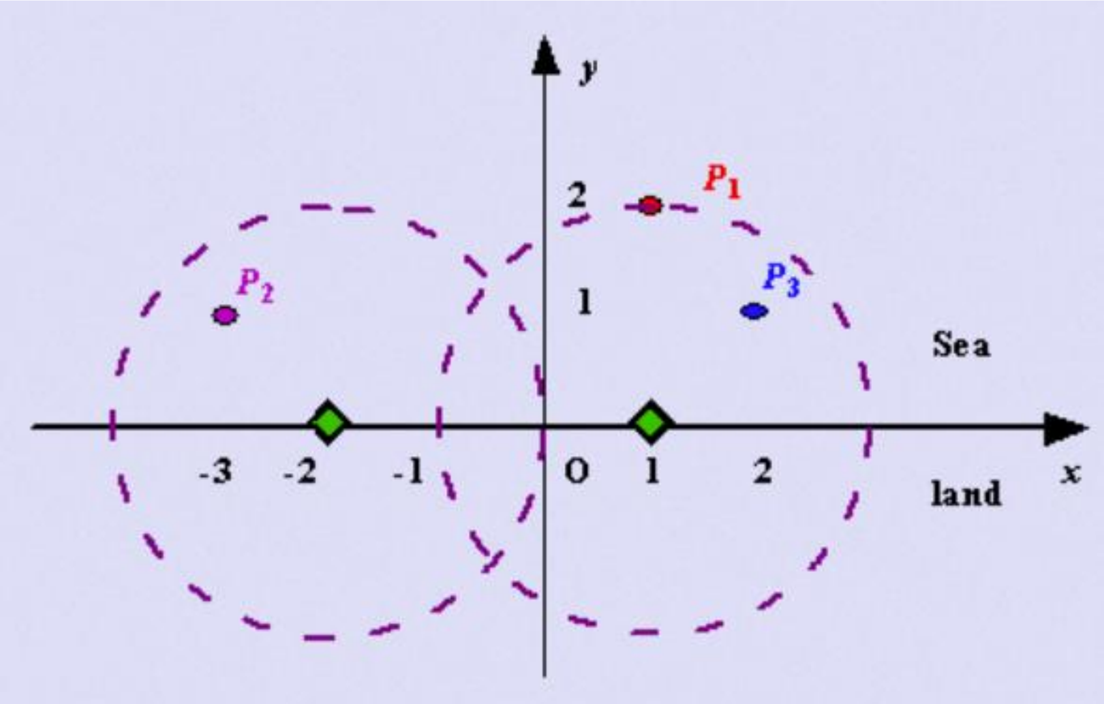
加强版： SJOJ1991 考试前一天有n个同学来问问题，每个预约时间范围 $[s_j, d_j]$，助教只有m个，每个工作时长必须相等。
求： 助教最短工作时长

区间 | 区间插点

POJ1328 近海有n个岛，在海岸线上装雷达来监测，雷达半径d，每个岛坐标 (x_j, y_j) 求最少雷达数.
输入： n,d ; x1,y1,x2,y2,……，xn,yn;
输出： 最少的雷达数

问题转化： 给n个一维闭区间，找出最少的点，保证每个区间至少有一个点

注： 解题代码稍长，可以留着以后写



https://blog.csdn.net/iteye_18955/article/details/82046401?utm_medium=distribute.pc_relevant.none-task-blog-2%7Edefault%7EBlogCommendFromMachineLearnPai2-1.baidujs&depth_1-utm_source=distribute.pc_relevant.none-task-blog-2%7Edefault%7EBlogCommendFromMachineLearnPai2%7Edefault-1.baidujs

赶工 | 最小权最小延迟调度

变式1:有n个作业，每个需t_j天完成。第j个没按时完成每天扣 c_j分，截止日期d_j。
安排顺序最小化总扣分.
输入： n ; t1,t2,……，tn; c1,c2,……，cn; d1,d2,……，dn
输出： 最小的总扣分

变式2:有n个商品，每个需t_j天卖出。第j个按时卖出赚 p_j 钱，保质期d_j。
安排顺序最大化总收入.
输入： n ; t1,t2,……，tn; p1,p2,……，pn; d1,d2,……，dn
输出： 最大的总收入

赶工 | 最小延迟调度

问题：有n个作业，第j个需要 t_j 天完成，截止日期d_j，安排顺序最小化总延迟.

注：如果作业j在f_j天完成，则延迟为 l_j = max{0,f_j − d_j }, 最小化$L = \sum_j l_j$

输入：n ; t1,t2,……，tn; d1,d2,……，dn

输出： 最小的总延迟L

赶工 | 最小权延迟调度

HDU1789:有n个作业，每个需1天完成。第j个没按时完成扣 c_j分，截止日期d_j。安排顺序最小化总扣分.

输入： n ; c1,c2,……， cn; d1,d2,……， dn

输出： 最小的总扣分

POJ1456:有n个商品，每个需1天卖出。第j个按时卖出赚 p_j 钱，保质期d_j。安排顺序最大化总收入.

输入： n ; p1,p2,……， pn; d1,d2,……， dn

输出： 最大的总收入

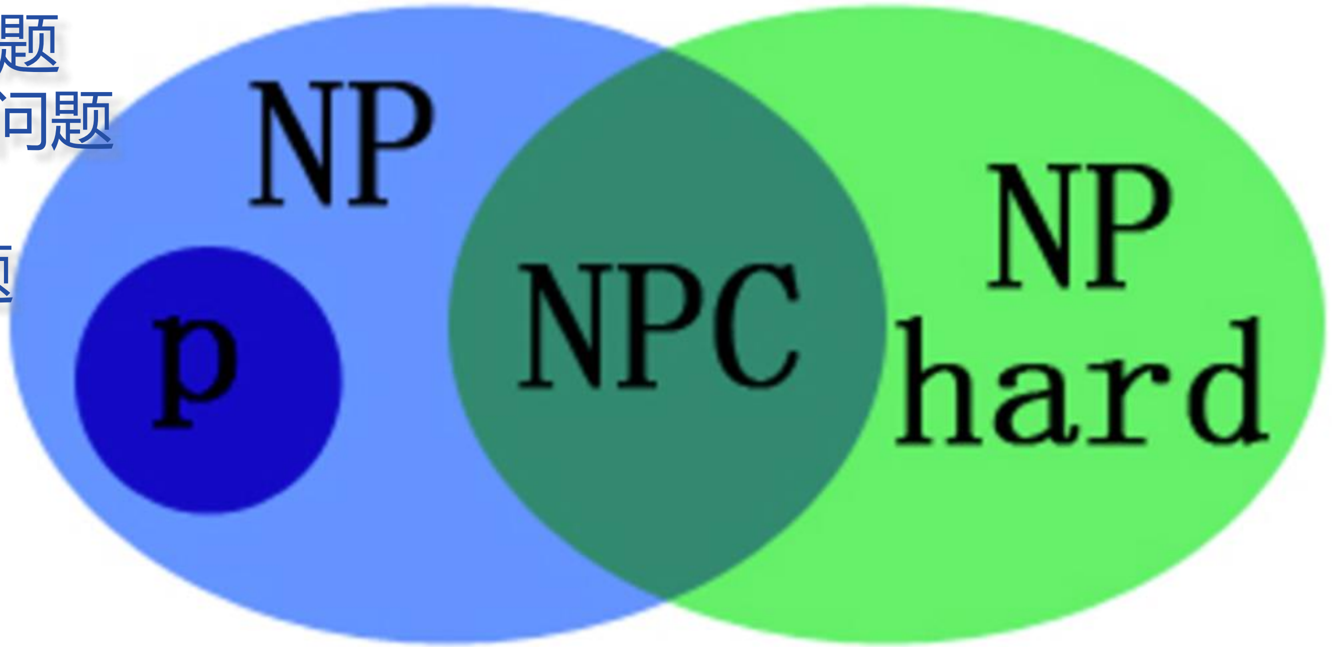
演示：优先队列，多输入

CONTENTS

计算复杂性——NP相关概念

有一些难度夸张问题：停机问题，打印全部排列
计算问题的复杂性：有P, NP, NP完全, NP难 ,PSPACE这些难度限制级
容易搞不清的是NP,NP完全(NPC) 和NP难(NPH)这三个概念

P类问题：存在多项式时间求解算法的问题
NP类问题：存在多项式时间验证算法的问题
如旅行商问题的判定形式
NP难问题：“通吃”所有NP问题的问题
如逻辑电路问题
NP完全问题：是NP问题；是NP难问题
逻辑电路问题
0-1整数规划
旅行商问题
……



https://blog.csdn.net/weixin_44810016/article/details/103394067

https://blog.csdn.net/qq_27328865/article/details/89468384

计算复杂性——NP完全问题

新千年世界七大数学难题

1.P=NP?

2.霍奇猜想

3.庞加莱猜想 √2006年8月,佩雷尔曼

4.黎曼假设 疑似解决

5.杨-米尔斯存在性和质量缺口

6.N-S方程的存在性与光滑性

7.BSD猜想

NP完全问题举例：旅行商问题(TSP)

给定n个城市和两两距离，问是否存在一条长度不大于D的环游路线

计算复杂性——NP完全问题举例

图论组合优化问题

简称	全称	中文名
TSP	Traveling Salesman Problem	旅行商问题
VRP	Vehicle Routing Problem	车辆路径问题
MaxCut	Maximum Cut	最大割问题
MVC	Minimum Vertex Cover	最小顶点覆盖问题
MC	Maximum Clique	最大团问题
MIS	Maximum Independent Set	最大独立集问题
MCP	Maximum Coverage Problem	最大覆盖问题
MDS	Minimum Dominating Set	最小支配集问题
GC	Graph Coloring	图着色问题
GM	Graph Matching	图匹配问题

组合优化问题

简称	全称	中文名
SAT	Boolean Satisfiability	可满足性问题
BP	Bin Packing	装箱问题
KP	Knapsack Problem	背包问题
JSP	Job Scheduling Problem	任务调度问题
IP	Integer Programming	整数规划问题

背包问题：给定n个物品的价值v_i和重量w_i,背包容量W，求最大价值

动态规划算法复杂度O(nW)

多项式算法？ 伪多项式算法！

近似算法——三种求解NP完全问题的路线之一

NP完全问题的含义——三者不可兼得

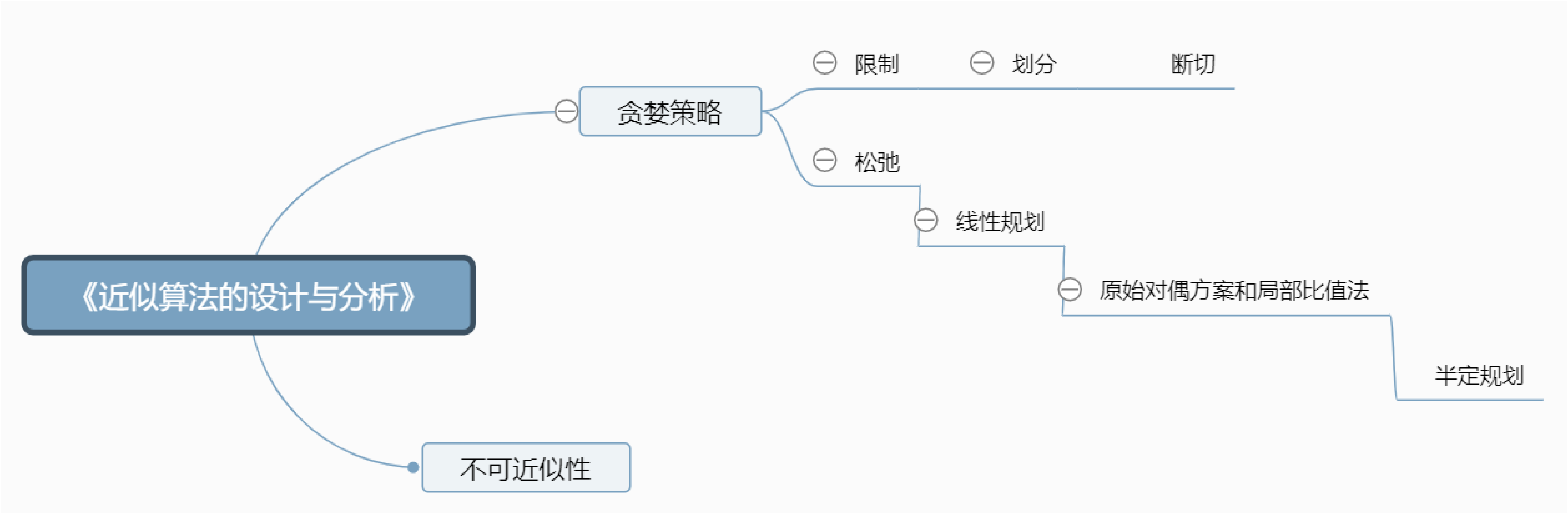
- 算出问题的最优解.
- 算出大规模问题的解（或者说多项式时间算法）.
- 算出任意形式问题的解。

NP完全问题的解法

- 不要最优解——近似算法，启发式搜索.
- 不要大规模——指数级搜索.
- 不要任意形式—— 求解特殊形式(如二分图最小点覆盖)

近似算法思路： 贪心法、线性规划舍入、划分……

近似算法设计策略



《近似算法的设计与分析》(堵丁柱 葛可一 胡晓东)

近似算法——贪心

三个例题

负载均衡：n个作业，每个处理时间t_j,分配给m个机器，使得负载最大的机器工作时间最短

中心选址： n个客户地点，选择k个中心，使得每个客户到最近中心的最大距离最短

集合覆盖：给定全集U 和n个子集S_j, 选择每个子集的代价为w_i，是否可以选择总权重不超过k的S_j，使$\bigcup_{j=1}^{k} S_j$ =U

贪心近似算法

负载均衡：最短分配，近似比2
排序后最短分配，近似比1.5

中心选址： 对中心半径r进行二分搜索，以客户点为中心，每次加最远点，近似比2

集合覆盖：贪心选择$\frac{w_i}{|S_i \cap R|}$的子集，其中R为全集中尚未被覆盖的元素。
近似比：$H(\max|S_i|) = \sum_{i=1}^{\max|S_i|} \frac{1}{i}$

近似算法——不可近似性

旅行商问题有近似算法

- **贪心1最近邻算法(类似Prim)**
  - 在每次没到过的点中选最近的
  - 近似比$1+\frac{\log n}{2}$
- **贪心2(类似Kruscal)**
  - 每次连接距离最近的两个不连通的点
  - 近似比$\frac{1}{2}+\frac{\log n}{2}$
- **插入算法 Insertion Algorithm**
  - 选择一个节点，插入环路中
  - 最小 cheapest / 最近 nearest / 最远 fastest / 随机 random

TSP不存在多项式时间α-近似算法

否则NPC问题哈密顿回路问题就有多项式解

除非P=NP

若假设距离**满足三角不等式** |AB|+|AC|≥|AC|

最小生成树算法：近似比2

Christofides 算法：近似比1.5

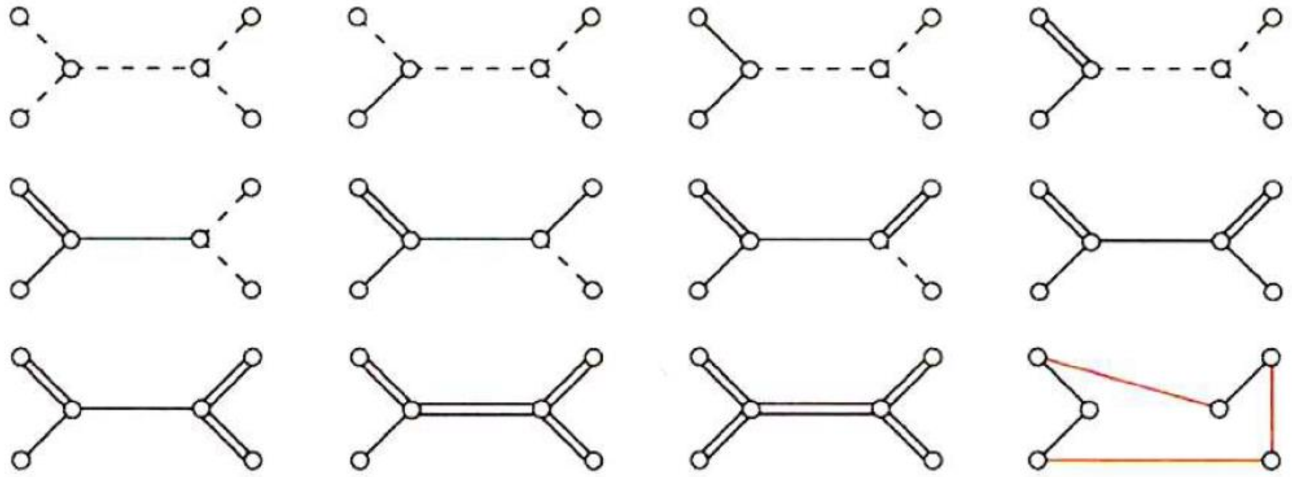
复杂度$O(n^2\log n)$

TSP的不可近似比是123/122

TSP不存在常数近似比算法的证明：《近似算法的设计与分析》(堵丁柱 葛可一 胡晓东) P26

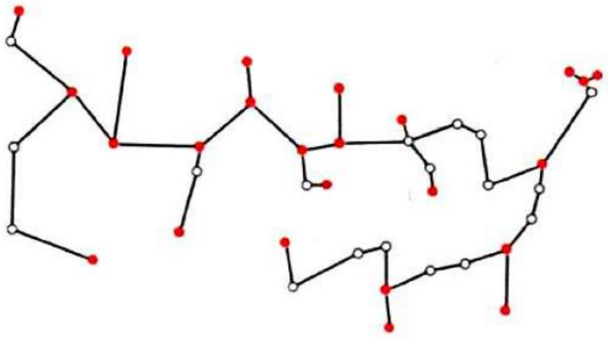
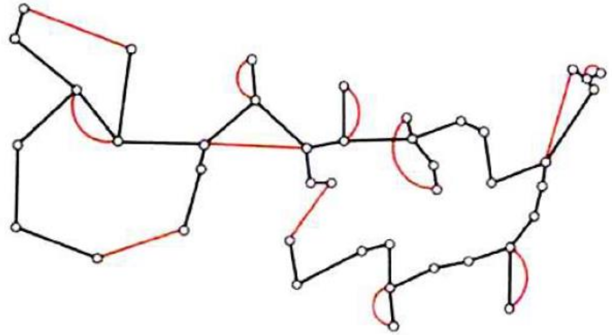
近似算法——TSP

- 最小生成树算法
  - Kruskal/Prim算法得到最小生成树
  - DFS这棵树，访问点的顺序即为TSP路径
  - 2近似比：算法路径长≤2倍最小生成树长度≤2倍TSP最优解去掉任何一条边



近似算法——TSP

- Christofides Algorithm
  - 最小生成树T
  - T中度数为奇数的顶点集O，O有偶数个顶点
  - 构造O在原完全图上的最小完全匹配M
  - M和T合并，得到H，其中每个顶点都是偶数度的
  - H可以形成一个欧拉回路
  - 将欧拉回路改造为哈密尔顿回路：跳过重复点
  - 近似比：3/2，时间复杂度：$O(n^2 \log n)$

CONTENTS

贪心算法理论初步——次模函数、独立系统、拟阵

离散数学概念

1. 次模函数

2. 独立系统、拟阵

贪心算法结果

1. 单调非负次模函数$f(S)$最大化问题的贪心算法近似比$1-\frac{1}{e}$

2. 证明最小生成树、任务调度问题(最小超时惩罚)贪心算法的最优性

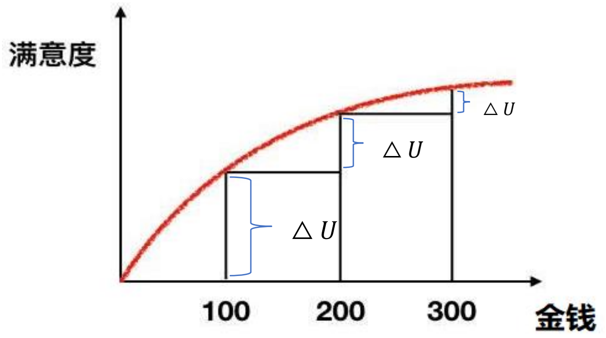
次模优化 https://alan20210202.github.io/2020/07/06/PACT0704/
拟阵 https://www.luogu.com.cn/blog/cpp/ni-zhen-yu-zui-you-hua-wen-ti

边际效用

- “效用” = “效益” = “满意度”。
- 边际效用：新增加的最后一单位的“消费”所获得的“效用”（“满意度”）。
- 边际效用递减规律：同一享乐不断重复，则其带来的享受逐渐递减。
  - 当消费者消费某一物品的总数量越来越多时，其新增加的最后一单位物品的消费所获得的效用（即边际效用）通常会呈现越来越少（递减）的现象。
  - 开始的时候，收益值很高，越到后来，收益值就越少。
    - x是自变量，U是因变量，U随x的变化而变化，随着x值的增加，U的增加量在不断减小。
    - 特点：非线性关系；$\frac{\triangle U}{\triangle x}$ 逐渐变小
    - 例子：身无分文的人获得100元时的满意度

vs

身缠万贯的人获得100元时的满意度



次模函数

- 次模函数（submodular function）又称“子模函数”或“亚模函数”。
- 次模函数具有次模性（submodularity，即，边际效应一直在减少）。
  - 它是经济学上边际效用递减（property of diminishing returns）现象的形式化描述。
- 给定一个集合函数 f: $2^V \to \mathbb{R}$，其将有限集 V 的一个子集 $S \subseteq V$ 映射为一个实数。
  - 定义1：如果对于 V 的任意两个子集 S、T，满足：

$$f(S \cup T) + f(S \cap T) \leq f(S) + f(T)$$

  则称 $f(\cdot)$ 是次模函数。
  - 定义2：对任意的$S \subseteq T \subseteq V$，并且 $e \in V \setminus T$，满足：

$$f(T \cup \{e\}) - f(T) \leq f(S \cup \{e\}) - f(S)$$

  则称 $f(\cdot)$ 是次模函数。

次模函数常用性质

- 在计算机领域，次模函数的常用性质：
  - 性质1. 下面这类“次模函数最大化”问题是NP-hard的：
    - 对于一个次模函数f，如果给定一个限制条件C，要求找出一个满足条件C的集合S，使得$f(S)$最大。
  - 性质2. 对于上述NP-hard问题，可以套贪心模板进行求解（多项式时间）：
    - 贪心思路：每次迭代地在解中加入增益最大且满足条件C的元素
    - 在第i次迭代时，解为$S_i = S_{i-1} \cup \{argmax_e \Delta(e|S_{i-1})\}$，其中$\Delta(e|S_{i-1}) = f(S_{i-1} \cup \{e\}) - f(S_{i-1})$
  - **性质3.** 如果一个次模函数f是单调且非负的，即对于任意$e \notin S$，$f(S \cup \{e\}) \geq f(S)$且对于任意$S \subseteq V$，$f(S) \geq 0$，那么上面的贪心算法找出的近似解的近似比为$\boxed{(1 - \frac{1}{e})}$。
- 难点：证明（构造）求解问题的目标函数是一个（单调非负）次模函数。

参考资料

- 《算法导论》(Cormen)
- 《算法设计》(Kleinberg)
- 《近似算法的设计与分析》（堵丁柱)
- CS214算法课(高晓沨)https://anl.sjtu.edu.cn/gao-xf/course/CS214-2021
- CS222算法课(蒋力)资料
- 2020年12月23日 算法组小组会 卢昊桢
- 2020年7月19日 机器学习求解组合优化 吴关昊
- 怎么理解次模函数 submodular function? https://www.zhihu.com/question/34720027
- Nemhauser, George L., Laurence A. Wolsey, and Marshall L. Fisher. "An analysis of approximations for maximizing submodular set functions—I."Mathematical Programming 14.1 (1978): 265-294.

谢谢！