计算复杂性简介

引子

- 你是某大公司的骨干，经常得到大老板的赏识。
- 某天大老板给你个简单的问题，很快你就给出了答案。
- 不久，大老板给你同样的问题，只不过这次数据变大了，你满怀信心的说“So easy!”。
- 一小时过去了，一天过去了，一个月过去了，一年过去了……结果……？？？

集合划分问题

- 给定一个数集A，问是否可以将其分成两个不相交的子集，使得两个子集中的元素之和相等？
- 若所有元素之和为奇数，则显然不行。但是，元素之和为偶数呢？
- {1, 3, 6, 8}

  1+8=3+6=9
- {2, 7, 15, 28, 34, 48, 56, 77, 106, 113}

  2+7+15+106+113=28+34+48+56+77=243

- {1, 5, 9, 16, 18, 23, 28, 29, 35, 37, 39, 41, 45, 48, 49, 55, 56, 67, 68, 69, 72, 75, 78, 79, 81, 86, 88, 95, 100, 106, 112, 115, 118, 122, 125, 127, 135, 138, 140, 145, 149, 155, 162, 168, 177, 178, 180, 195, 199, 201, 205, 209, 215, 220, 224, 228, 229, 233, 238, 239, 245, 248}共62个数

1+5+……+168+177=178+180+……+245+248=3686

枚举法（穷举法）

- 集合总共有 $\sum_{i=0}^{n}\frac{1}{2}\binom{n}{i}$ 种不同分法。
- 假定检查和计算每种分法是否可行只需1微秒（10^6微秒=1秒）。
- 当n=50，约需35.7年！
- 当n=60，约需366个世纪！！！

旅行售货商问题（TSP）
The Traveling Salesman Problem

- 一个商人从某个城市出发，到其他城市去售卖商品，他希望在旅途中恰好路过每个城市一次，最后返回出发地。如何安排旅行线路，使得所走路线总长度最短。
- 枚举法：找出所有的可行路线，然后比较找出最短的。
- N个城市，可行的路线有(N-1)!/2 .
- 计算一个可行路线只需1纳秒（10^9纳秒=1秒）。
- 遍历23个城市大概需要 178个世纪！！！

若干 定义

- 定义1：优化问题π是一个极小化（极大化）问题，它由下述三部分组成：

（1）实例集合；

（2）对每一个实例I，有一个有穷的可行解集合$S(I)$；

（3）目标函数f，它对每一个实例I和每一个可行解$\sigma \in S(I)$，赋予一个有理函数$f(I,\sigma)$.

- 定义2：算法是指一步步求解问题的通用程序，它是解决问题的程序步骤的一个清晰描述.

- 定义3：对于一个优化问题，如果给定任意一个实例，算法总能找到一个可行解，那么就称之为近似算法；如果进一步这个可行解的目标值总等于最优值，则称之为最优算法.
- 问：是否任何数学问题都有算法求解呢？
- 答案是否定的.
- 停机问题：是否存在一个算法，对于任意给定的图灵机都能判定任意的初始格局是否会导致停机？
- 著名英国数学家图灵（Turing）证明了不存在一个算法，它能对该问题的一切实例给出正确答案.

- 定义4：算法时间复杂性是关于实例输入长度n的函数$f(n)$，用来表示算法的时间需求. 对于每一个可能的输入长度，它是该算法解此输入长度的最坏可能的实例所需的时间（基本运算步数）。
- 定义5：若存在一个常数C，使得对所有$n \geq 0$，都有$|f(n)| \leq C|g(n)|$，则称函数$f(n)$是$O(g(n))$. 时间复杂性是$O(p(n))$的算法称为多项式时间算法. 不能这样限制时间复杂性函数的算法称为指数时间算法.

$$O(n\log n), O(n^{3.7}), O(n^2), O(n^{\log n}), O(n!), O(2^n)$$

几个多项式和指数时间复杂性函数增长情况			
函数	规模n的近似值		
	10	100	1000
n	10	100	1000
$n\log n$	33	664	9966
n^3	10^3	10^6	10^9
2^n	1024	$1.27*10^{30}$	$1.05*10^{301}$
$n!$	3628800	10^{158}	$4*10^{2567}$

1小时内可解的问题实例的规模			
函数	用现在计算机	用快10倍的计算机	用快100倍的计算机
n	N_1	$10N_1$	$100N_1$
$n\log n$	N_2	$8.2N_2$	$67N_2$
n^3	N_3	$2.15N_3$	$4.64N_3$
2^n	N_4	$N_4+3.32$	$N_4+6.64$
$n!$	N_5	$\le N_5+2$	$\le N_5+3$

	10	20	30	40	50	60
n	.00001	.00002	.00003	.00004	.00005	.00006
n^2	.0001	.0004	.0009	.0016	.0025	.0036
n^3	.001	.008	.027	.064	.125	.216
n^5	.1	3.2	24.3	1.7分	5.2分	13分
2^n	.001	1.0	17.9分	12.7天	35.7年	366世纪
3^n	.059	58分	6.5年	3855世纪	2×10^8 世纪	1.3×10^{13} 世纪

- 定义6：一个优化问题如果已经找到了多项式时间算法，那么就称它为多项式时间可解问题. 把所有这样的问题集合记为P，因此多项式时间可解问题就称为P类问题.
- 定义7：答案为“是”或“否”的问题称为判定问题.
- 问：优化问题与其相应的判定问题的困难程度是否相当？