第一章 NP 完全性理论

1.1 概念

定义 1.1 (确定性算法与非确定性算法). • **确定性算法**: *算法中每个操作结果唯一确定, 算法操作结果也是唯一确定.*

• **非确定性算法**: *算法操作结果不唯一，而是来自可能值集合.*

这里的确定性算法指的就是非随机算法，非确定性算法就是随机算法，可能值集合也就是可行解.

定义 1.2 (判定问题，判定算法). *答案非 0 即 1 的所有问题称为判定问题，求解判定问题的算法称为判定算法.*

往往将可在多项式时间内解决的问题视为“易”解问题，需要指数时间及以上解决的问题称为“难”解问题. 为了详细区分这两种问题，引入了 P 类问题和 NP 类问题，首先要注意，NP 问题不是指不能在多项式时间内求解的问题，而是值**无法确定**是否可在多项式时间内求解的问题，因为可能通过随机算法，运气好直接碰出答案了，所以需要用算法的确定性对其进行定义.

定义 1.3 (P 类问题). *在多项式时间内，可使用确定性算法求解的判定问题构成的集合.*

定义 1.4 (NP 类问题). *在多项式时间内，可使用非确定性算法求解的判定问题构成的集合.*

在默认使用的均是确定性算法前提下，可以 P 类问题与 NP 类问题可以如下表示：

$$
\begin{aligned}
\text{P 类问题} &= \{\text{判定问题} : \text{可在多项式时间内}\textbf{求出}\text{正确解}\}\\
\text{NP 类问题} &= \{\text{判定问题} : \text{无法确定是否可在多项式时间内求出正确解}\}\\
&= \{\text{问题} : \text{可在多项式的时间内}\textbf{验证}\text{是否得出正确解}\}
\end{aligned}
$$

可以从上述 NP 类问题的第二种定义可以看出，将问题转化为“判定问题”是必要的，使得验证时间为 $\mathcal{O}(1)$，仅需关心求解所需的时间复杂度. 证明一个问题是否是 NP 问题，只需判断**是否可以在多项式时间内验证是否为正确解**.

由定义可知，P $\subset$ NP，但 P 是否等于 NP 是世纪难题.

在引入 NP 完全问题和 NP 难定义前，先给出这些定义之间的逻辑关系：由于没有定义 NP 问题就是不能在多项式时间内求解的问题，所以还是可能存在一种线性算法使其能在多项式时间内给出正确解，只是人们还没有发现（如果 P = NP 则说明全部的 NP 问题都能多项式时间内解决，而且这种“可能”现在有可能可以通过量子计算机完成）. 对于 NP 问题 A，通过“多项式时间变换”，将 NP 问题 A 变换为更为复杂的 NP 问题 B，如果可以多项式时间内求解 NP 问题 B，则可以求解 NP 问题 A，人们将比所有 NP 问题都更难的那部分 NP 问题称为 NP 完全问题，并将比所有 NP 都要难的问题通称为 NP 难问题.

定义 1.5 (多项式时间变换). *设 X, Y 分别是定义在实例集 I, J 上的两个判定问题，若存在从 I 到 J 的映射 $f : I \to J$ 使得*

1. *f 是在多项式时间内可计算的映射.*
2. $\{f(x) : x \in X\} \subset Y$

则称问题 X 能**多项式时间变换**为问题 Y，记为 $X \propto_p Y$.

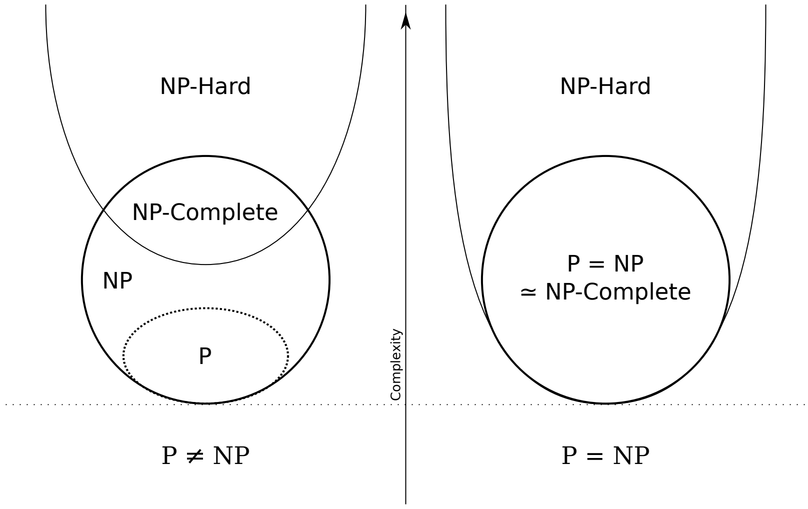
定义 1.6 (NP 完全问题，NPC)**.** 设 $Y \in NP$，则 Y 是 *NP* 完全的，当且仅当，

1. $Y \in NP$.
2. $\forall X \in NP$ 有 $X \propto_p Y$.

将全体 *NP* 完全问题，记为 *NP* 完全类，简写为 *NPC*.

定义 1.7 (NP 难问题，NP-Hard)**.** 设 $Y \in NP$，则 Y 是 *NP* 难的，当且仅当，$\forall X \in NP$ 有 $X \propto_p Y$.（即满足 *NP* 完全问题中的第二条即可）

P, NP, NPC, NP-Hard 问题关系如下图[1] 所示，纵轴显示了解决问题的难度：



1.2 判断一个问题是否是 NP 完全问题

首先要记住几个经典的 NP 完全问题（重要的问题会在文末详细给出）：布尔表达式可满足问题（SAT），合取范式的可满足问题（CNF-SAT），三元合取范式可满足问题（3-SAT），k 团问题（CLIQUE），k 顶点覆盖问题（VERTEX-COVER），子集和问题（SUBSET-SUM），哈密顿回路问题（HAM-CYCLE），旅行商问题（TSP）.

Cook 定理说明 SAT 是 NP 完全问题，上述相邻的两个问题，前一个问题可通过多项式时间变换转化为后一个问题.

证明一个问题 A 是否是 NP 完全问题：假设已知问题 B 是 NP 完全问题，根据定义，只需证 A 是 NP 问题且 A 是 B 的 NP 完全问题即可.

第一步：证明 A 是 NP 问题，即是否可以在多项式时间内给出**验证**是否是问题 A 的正确解.

第二步：找一个多项式时间内的映射，将问题 B 映射到问题 A 上即可.

若只需证明 A 是 NP 难问题，则只需证明第二步.

k 团问题（CLIQUE）：给定一个无向图 $G = (V, E)$ 和一个正整数 k，判定图 G 是否包含一个大小为 k 的团，即是否存在，$V' \subset V$, $|V'| = k$ 且 $\forall u, w \in V'$ 有 $(u, w) \in E$.

k 顶点覆盖问题（VERTEX-COVER）：给定一个无向图 $G = (V, E)$ 和一个正整数 k，判定是否存在 $V' \subset V, |V'| = k$，使得 $\forall (u, v) \in E$ 有 $u \in V'$ 或 $v \in V'$. 若存在这样的 V'，则称 V' 为图 G 的一个大小为 k 顶点覆盖.

[1] By Behnam Esfahbod, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=3532181

子集和问题（SUBSET-SUM）：给定整数集合 S 和一个整数 t，判定是否存在 S 的一个子集 S' $\subset S$，使得 S' 中整数的和为 t.

哈密顿回路问题（HAM-CYCLE）：给定无向图 $G = (V, E)$，判定其是否含有哈密顿回路，即判定 G 是否存在经过 V 中各顶点恰好一次的回路.

旅行商问题（TSP）：给定一个无向完全图 $G = (V, E)$ 及定义在 $V \times V$ 上的一个费用函数 c 和一个整数 k，判定 G 是否存在经过 V 中各顶点恰好一次的回路，使得该回路的费用不超过 k.