Computational Complexity

Lecture 13: Average-case complexity and Impagliazzo's Five Worlds

Ronald de Haan
me@ronalddehaan.eu

University of Amsterdam

March 15, 2021

Recap

- Subexponential-time complexity
- Exponential-Time Hypothesis (ETH)

What will we do today?

- Average-case complexity
- One-way functions
- Impagliazzo's Five Worlds

Worst-case complexity

- A problem $L \subseteq \{0,1\}^*$ can be solved in *worst-case running time* $T(n)$ if there exists an algorithm A that solves L and that halts within time $T(|x|)$ for each $x \in \{0,1\}^*$.

- In other words, the worst-case running time $T(n)$ is the maximum of the running times for all inputs of size n.

Distributional problems

Definition (distributional problems)

A *distributional problem* $\langle L, \mathcal{D} \rangle$ consists of a language $L \subseteq \{0,1\}^*$ and a sequence $\mathcal{D} = \{\mathcal{D}_n\}_{n \in \mathbb{N}}$ of probability distributions, where each $\mathcal{D}_n$ is a probability distribution over $\{0,1\}^n$.

The class distP / avgP

Definition (distP)

$\langle L, \mathcal{D} \rangle$ is in the class distP (also called: avgP) if there exists a deterministic TM $\mathbb{M}$ that decides L and a constant $\epsilon > 0$ such that for all $n \in \mathbb{N}$:

$$\mathop{\mathbb{E}}_{x \in_R \mathcal{D}_n} [\ \text{time}_{\mathbb{M}}(x)^{\epsilon}\ ] \text{ is } O(n).$$

- The ϵ is there for technical reasons—to invert a polynomial to $O(n)$.

Polynomial-time computable distributions

Definition (P-computable distributions)

A sequence $\mathcal{D} = \{\mathcal{D}_n\}_{n\in\mathbb{N}}$ of distributions is P-*computable* if there exists a polynomial-time TM that, given $x \in \{0,1\}^n$, computes:

$$\mu_{\mathcal{D}_n}(x) = \sum_{\substack{y\in\{0,1\}^n \\ y\leq x}} \mathbb{P}_{\mathcal{D}_n}[y],$$

where $y \leq x$ if the number represented by the binary string y is at most the number represented by the binary string x.

Polynomial-time samplable distributions

Definition (P-samplable distributions)

A sequence $\mathcal{D} = \{\mathcal{D}_n\}_{n\in\mathbb{N}}$ of distributions is P-*samplable* if there exists a polynomial-time probabilistic TM $\mathbb{M}$ such that for each $n \in \mathbb{N}$, the random variables $\mathbb{M}(1^n)$ and $\mathcal{D}_n$ are equally distributed.

The class distNP and sampNP

Definition (distNP)

A problem $\langle L, \mathcal{D} \rangle$ is in distNP if $L \in \mathsf{NP}$ and $\mathcal{D}$ is P-computable.

Definition (sampNP)

A problem $\langle L, \mathcal{D} \rangle$ is in sampNP if $L \in \mathsf{NP}$ and $\mathcal{D}$ is P-samplable.

- The questions "distNP $\stackrel{?}{=}$ distP" and "sampNP $\stackrel{?}{=}$ distP" are average-case analogues of the question "NP $\stackrel{?}{=}$ P"

One-way functions (OWFs)

Definition (one-way functions)

A polynomial-time computable function $f : \{0,1\}^* \to \{0,1\}^*$ is a *one-way function* if for every polynomial-time probabilistic TM $\mathbb{M}$ there is a neglegible function $\epsilon : \mathbb{N} \to [0,1]$ such that for every $n \in \mathbb{N}$:

$$\mathop{\mathbb{P}}_{\substack{x \in_{\mathbf{R}} \{0,1\}^n \\ y = f(x)}} \left[\ \mathbb{M}(y) = x' \text{ such that } f(x') = y\ \right] < \epsilon(n)$$

where a function $\epsilon : \mathbb{N} \to [0,1]$ is *neglegible* if $\epsilon(n) = \frac{1}{n^{\omega(1)}}$, that is, for every c and sufficiently large n, $\epsilon(n) < \frac{1}{n^c}$.

- Conjecture: there exist one-way functions (implying P $\neq$ NP)
- OWFs can be used to create private-key cryptography

Impagliazzo's Five Worlds (1995)

Five possible situations regarding the status of various complexity-theoretic assumptions:

- Algorithmica
- Heuristica
- Pessiland
- Minicrypt
- Cryptomania

Russell Impagliazzo. *A personal view of average-case complexity.* In: Proceedings of the 10th Annual IEEE Conference on Structure in Complexity Theory, pp. 134–147, 1995.

Algorithmica

- $P = NP$ (or $NP \subseteq BPP$)

- Say, SAT is linear-time solvable
- This is a computational utopia
- There exist efficient algorithms for creative tasks, e.g., writing proofs
- Essentially no cryptography possible (private-key nor public-key)

Heuristica

- $\mathsf{P} \neq \mathsf{NP}$, but $\mathsf{distNP}, \mathsf{sampNP} \subseteq \mathsf{distP}$

- Breakthroughs of $\mathsf{P} = \mathsf{NP}$ work almost all the time
- So cryptography breaks too

Pessiland

- $\mathsf{distNP}, \mathsf{sampNP} \not\subseteq \mathsf{distP}$ (so $\mathsf{P} \neq \mathsf{NP}$)
- one-way functions do not exist

► No computational breakthroughs, and most cryptography schemes do not work

Minicrypt

- One-way functions exist (so $\mathsf{P} \neq \mathsf{NP}$ and $\mathsf{distNP} \not\subseteq \mathsf{distP}$)

- No "$\mathsf{P} = \mathsf{NP}$"-type breakthroughs
- Private-key cryptography works
- All "highly structured" problems in NP, such as integer factoring, are solvable in polynomial-time
- Public-key cryptography might not work

Cryptomania

- Factoring large integers takes exponential time on average (or a corresponding result for a similar problem)

- ► No general-purpose efficient algorithms ($\mathsf{P} \neq \mathsf{NP}$)
- ► Private-key and public-key cryptography works

Impagliazzo's Five Worlds (1995)

- Five worlds:
  - Algorithmica – efficient general-purpose algorithms
  - Heuristica
  - Pessiland – worst of all worlds
  - Minicrypt
  - Cryptomania – all kinds of cryptography possible
- (Technically, these cases are not exhaustive—there are some "weirdland" scenarios, e.g., the case where SAT $\in$ P, but the fastest algorithm takes time $\Theta(n^{100})$.)

Recap

- Average-case complexity
- One-way functions
- Impagliazzo's Five Worlds

Next time

- Recap and/or question session