CS 229r Information Theory in Computer Science Mar 14, 2019

Lecture 14

Instructor: Madhu Sudan *Scribe: Amir Shanehsazzadeh*

1 Admin

- PSET 3 Due Tomorrow! (Last PSET)
- Project + *Team* Selection (Soft Deadline Tomorrow)
- (Likely) Will Post Practice Problems Tomorrow

2 Today

- Future/Project Topics

3 Information Complexity

Right now we are in the middle of Information Complexity and will explore deeper into this. So far we have defined the external complexity as

$$\mathrm{IC}_{ext}(\pi) = I(XY;\pi),$$

which is the information about XY given the protocol π as viewed by an external observer. We can also define internal IC as

$$\mathrm{IC}_{int}(\pi) = I(Y;\pi|X) + I(X;\pi|Y),$$

which is the information that Alice and Bob learn about X and Y from the protocol π.
The following inequality is not obvious for general RV's but it is for protocols π:

$$\mathrm{IC}_{int}(\pi) \leq \mathrm{IC}_{ext}(\pi) \leq \mathrm{CC}(\pi).$$

3.1 IC and Communication

It turns out that IC limits the ability to communicate. A paper by [Barak et al.] proved the following main theorem:

Theorem 1 (Barak et al.)**.** *If π communicates C bits and requires I bits of information, then it can be compressed to a $\tilde{O}(\sqrt{IC})$ bit protocol.*

Ideally, we would be able to compress to a $\tilde{O}(I)$ bit protocol, but then we could just keep compressing. However, this is still strong as it gives a protocol asymptotic to the geometric mean of C and I.

3.2 IC and Amortized Communication

A later work by [Braverman, Rao] gives another theorem:

Theorem 2 (Braverman, Rao)**.** *Amortized communication complexity $\approx$ information complexity. Take t copies of X and Y: $\{X_1, X_2, ..., X_t\}$ and $\{Y_1, Y_2, ..., Y_t\}$. The amount of communication needed to transmit $\pi(X_i, Y_i)$ for $i \in [t]$ is clearly in the interval $[tI, tC]$, but by this theorem we can replace this with the interval $[tI, \tilde{O}(tI)]$. In fact, we can replace this with the interval $[tI, tI(1 + o(1))]$.*

This tells us that we can squish the result after taking many, many copies. The result led to hopes that IC would give information about CC without amortization. This question was open until work by [Gamar-Kol-Raz] and [Braverman]:

Theorem 3 (GKR). *$\exists f$ such that $IC_{int}(f) = k$ and $CC(f) \geq 2^k$.*

Theorem 4 (Braverman). *$\forall f$, $CC(f) \leq 2^{IC(f)}$.*

Interestingly, in the past many American and Soviet information theorists discovered identical results but could not communicate because of language barriers. Now different computer scientists find identical or very similar results but in different fields.
This amortized result was also shown by Distributed Source-Coding [Ishwar-Ma].

Exercise 5. *Potential project idea: Read this paper and compare its results to what we showed in class.*

3.3 Limits of Amortization

What should t be with respect to n? I is independent from t so the problem is more tractable, but we have

$$\mathrm{IC}(f) = \inf_{\pi \text{ that compute } f} \{\mathrm{IC}(\pi)\}.$$

To find this we must look at a countable but not finite set.

3.3.1 Entropy and Channel Capacity

Both single-shot and amortized entropy are computable, via Huffman encoding and direct computation, in Poly time. Single-shot channel capacity could be NP-Complete and amortized channel capacity can be done in Poly time via convex programming.

3.3.2 Zero-Error Channel Capacity

One our our papers considers the case where the error is 0 instead of being close to 0 with high probability. Single-shot is believed to be NP-complete and amortized may not be in P and may not even be computable. Work by [Lovasz] discovered lower bounds for the Shannon capacity. Note that this problem is much more combinatoric whereas the case where error goes to 0 is information-theoretic. The zero-error channel coding problem can be defined in terms of common randomness generation.

Definition 6 (Common Randomness Generation). Given $(XY) \sim \mu$ (potentially correlated) give Alice $\{X_1, ..., X_t\}$ and Bob $\{Y_1, ..., Y_t\}$. Also give both of them some private randomness R_A and R_B and have them communicate $\gamma \cdot t$ times between each other. We want Alice and Bob to output $\{R_1, R_2, ..., R_{\rho t}\}$ where $R_i \sim$ Unif and $\gamma, \rho \in (0, 1]$.

This begs the question of whether or not we can define a protocol on μ with parameters (γ, ρ). Can we get $\gamma < \rho$? The case where $X = Y \sim$ Unif requires no work. The hard case is when X and Y are only slightly correlated.
Work on this problem goes back to [Gacs, Korner]. The first paper mentioned in our list is by [Witsenhausen]. This paper considers the case where Alice draws R_1 and Bob draws R_1' from the distribution for which $\mathbb{P}[00] = \frac{1}{2}$ and $\mathbb{P}[01] = \mathbb{P}[10] = \frac{1}{4}$. Another paper by [Ahslwede-Ciszar] consider that case where (R_1, R_1') are usually equal, but occasionally not. Other papers, including one by [Cuff-Liu-Verde], look at the ratio

$$\max_{\pi} \left\{ \frac{\mathrm{IC}_{ext}(\pi)}{\mathrm{IC}_{int}(\pi)} \right\},$$

which is very fundamental in this line of research.

3.4 Computability of IC

$\mathrm{IC}(f)$ is not necessarily computable but is computably-approximable (i.e. can compute an ε-approximation in Poly time). [Braverman] shows this result. In a similar sense, some papers (one by [Ghazi, Kamath, Sudan]) prove theorems that tell us if parameters (γ, ρ) work for the common randomness generation problem. Another problem that is not known to be computable but is computably-approximable is a Markov chain we looked at before where the states are $X_1 \sim \mathrm{Bern}(\delta)$ and $X_2 \sim \mathrm{Bern}(\frac{1}{2} - \delta)$ and the transition matrix is

$$\begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix}.$$

Exercise 7. *Come up with an idea for the final project (maybe relating to IC)!*

4 Next Topics

The next topics we will look at are broad (some even form their own field).

4.1 Streaming Algorithms

- Communication complexity lower bounds lead to streaming lower bounds.
- How can you capture the limits/bounds? IT!

4.2 Data Structures

- I have large amounts of data and want it stored efficiently in a way that is efficiently/reliably accessible.
- Rich literature and active area of research

4.3 Differential Privacy

- How private is a mechanism? (Defined probabilistically)
- IT-tools are used to measures this via bounds.
- IT-tools are used to design mechanisms and to analyze limits.

4.4 Learning/Statistics/Finance (Portfolio Optimization)

4.5 Complexity of Optimization

4.6 Extension Complexity

- Paper by [Yannakakis] proved that any linear-time solution to TSP would require exponentially many constraints and thus be exponential. Communication complexity lower bounds were used to get restricted impossibility (symmetric linear programs).
- Paper by [Fiorini et al.] used DISJ lower bounds to show general impossibility.
- Paper by [Braverman, Moltra] used information complexity results for lower bounds.

4.7 Hardness of Approximation

- "2-prover proof systems"
- Parallel Repetition
- Started in [Raz '94], which opened the eyes of computer scientists to information theory. The work was improved in [Holenstein].

4.8 Shearer's Lemma

Equality in the inequality implies that the object is a box. What about approximate equality? [Ellis et al.] showed that approximate equality implies that the structure is approximately a box.