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

Lecture 14

Instructor: Madhu Sudan *Scribe: Ben Chen*

1 Overview

- We wrapped up our discussion of communication complexity of set disjointness last time
- This class will mainly be a discussion of various interesting problems/papers that are possible project topics
- No new topics before break
- I may end up presenting a few of these topics in class, but nothing is set in stone yet
- Links to the papers will be posted in the notes on the course website

1.1 Admin

- PSet 3 Due Friday
- Project and Team selection (soft deadline Friday)
- (Likely) will post practice problem Friday

2 Information Complexity

We defined information complexity as

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

which is the *external* information complexity, which is the complexity of the information from the point of view of an external observer who does not know either X or Y. There is also the *internal* information complexity, as Alice already knows X and Bob already knows Y, so we condition on knowing X and Y:

$$IC_{int}(\Pi) = I(Y;\Pi|X) + I(X;\Pi|Y).$$

The internal information complexity describes how much information the parties reveal about their inputs to each other by communicating, while the external information complexity describes how much information they would be revealing to a third party who can only see the transcript of the protocol.

We have that

$$IC_{int}(\Pi) \leq IC_{ext}(\Pi) \leq CC(\Pi).$$

The motivation for defining this is that it more closely represents what is going on in this scenario in some sense. An important paper relating to this is [Barak et al.]. In this paper, the authors proved:

Theorem 1. *If Π communicates C bits and reveals I bits of information about the inputs, then it can be compressed to a $\tilde{O}(\sqrt{IC})$ bit communication protocol.*

In words, this means that if players are typicallly sending empty bits every unit of time and only occasionally sending useful information, then we can compress the protocol quite a bit, although not, unfortunately, all the way to the entropy bound.

2.1 Amortized Communication Complexity

A later work of a few of the same authors [Braverman and Rao] gives

Theorem 2. *Amortized communication complexity ≈ Internal information complexity. That is, if we draw t samples $x_1, \ldots, x_t$ and $y_1, \ldots, y_t$ from X and Y, the amount of communication needed to transmit $\pi(x_i, y_i)$ is between tI and $tI(1 + o(1))$.*

Amortized communication complexity analogous to how we defined entropy: as the average over repeated trials, as opposed to the complexity of a single instance.

In this case, the theorem says we can almost compress a protocol to close to the internal information complexity in the amortized limit. This leads to the natural question of how communication complexity relates to information complexity in the non-amortized case.

Theorem 3 (Ganor-Kol-Raz)**.** *There exists f such that $IC_{int}(f) = k$, but $CC(f) \geq 2^k$.*

The complementary theorem by Braverman gave that

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

Papers that cover some version of these 2 theorems may or may not be covered in the rest of class and would make good final projects.

3 Information Theory and other CS

Distributed source coding [Ishwar-ma] seems to have preempted the amortized communication complexity by a few years.

[Ghazi-Jayram] finally drew from both CS and information theory papers.

Could be a good project to compare papers from information theory and CS about similar topics, since information theory was not at the forefront of CS until recently (many CS people did not know much about information theory, so many repeat papers exist, similar to the case of Russian/English repeat papers that are still being translated).

4 Computability of Information Complexity

The amortized complexity is much harder to compute since we need to consider the best algorithms on average, instead of simply focusing on a single instance, which makes for a more difficult analysis.

$$IC(f) = \inf_{\Pi \text{ that compute } f} \{IC(\Pi)\}$$

This "reduces" the exponential problem to a countable problem, which seems counterintuitive; how do we determine the best protocol among countably infinite possible protocols?

Entropy: Both single-shot and amortized are polytime. We can use Huffman coding for single-shot, which is computable efficiently. The amortized result is just the entropy, which is also computable efficienty.

Channel complexity: Here, the single-shot problem could already be NP-complete. The amortized problem, however, can be polytime using techniques like convex programming.

5 Zero-Error Shannon Capacity

What happens when we replace our exponential error bound with a requirement that our error actually vanishes? It is known that the single-shot zero-error channel capacity problem is NP-complete. However, we do not even know whether the amortized problem is computable or in P.

Note that this is the Shannon capacity of the graph. [Lovász] came up with a beautiful way to study the lower bounds on the Shannon capacity. He also introduced semi-definite programming. Note that the zero-error and error going to zero contexts are quite different; the former is combinatorial in flavor, the latter is more information theoretic.

6 Common Randomness Generation

One motivation for information complexity was the following problem:

Common Randomness Generation: We have $(X, Y) \sim \mu$. Alice draws from X and Bob from Y, say $x_1, \ldots, x_t$, $y_1, \ldots, y_t$. After some process, they each want to generate uniformly random bits

$$R_1, \ldots, R_{\rho t} \text{ and } R'_1, \ldots, R'_{\rho t}$$

Does μ permit protocols with (γ, ρ), where ρ is the common random bits per sample, and γt is the communication? An obvious approach would be to ignore the other distribution; we are interested in reducing $\gamma < \rho$.

There is rich literature about this problem, including a paper by [Witsenhausen].

Maybe we can first transform the type of randomness we have with a simple process. Consider, for example, (XY) is 00 with 1/2, 01 with 1/4, and 10 with 1/4. Can we output R_1, R'_1, which are usually equal, occasionally not?

A related topic to consider is also Secret Key Generation.

[Ahlswede-Csizar] continued the study of this problem. [Cuff-Liu-Verdu] asked about the ratio between the internal and external information

$$\max_{\Pi} \left\{ \frac{IC_{ext}(\Pi))}{IC_{int}(\Pi)} \right\}$$

This ratio turns out to be very fundamental in Common Randomness Generation and related problems and is worth looking into.

Note that when we were talking about $IC(f)$, it may not be actually be computable (since it may be some random transcendental value which may not be computable), but it is computably-approximable, that is, we can find an additive ε approximation, which is discussed in [Braverman].

[Ghazi,Kamath,S] (γ, ρ) is computably-approximable. This is the weakest possible result that still deserves the name "result." We don't have the time-complexity of this at all; it could be very suepr-exponential.

Some things do not have computable-approximability, such as the HMM thath transitions between 1 states, one that generates 0,1 uniformly at random, the other that emits 0.

7 Connections and Applications

7.1 Streaming

Another very important connection of information complexity is to streaming algorithms. Streaming algorithms deal with a very large stream of data you are receiving that you must operate on with both limited time and space. Comm complexity bounds lead to streaming lower bounds.

7.2 Data Structures

Best way to store data with low space after preprocessing, so that certain queries about the data can be answered quickly. This is related to streaming algorthims.

7.3 Differential Privacy

Mechanisms for ensuring privacy of users. Information theory is used both to design algorithms and analyze limits. There is a very vast literature on this topic.

There are also applications to learning, statistics, and finance. Some are more precisely mathematical, while others are more heuristic; both are appropriate for projects.

7.4 Optimization

Information theory has also been able to help us with the complexity of optimization. An interesting subarea is that of Extension Complexity, started in the late 1980s by [Yannakakis] as a response to a crank trying to prove $P = NP$ using linear algorithms. Yannakakis showed that all possible such approaches are hopeless by showing that the travelling-salesman polytope requires an exponential description in high dimensions using lower bounds from communication complexity. Yannakakis restricted himself to symmetric linear programs; the more general result was shown by [Fior et al] using disjointness lower bounds about 8 years later. THe work is continued by [Braverman-Mitra].

On the other hand, we can also study how certain optimization problems cannot be solved approximately very well. The central theme of showing hardness of approximation from the 1990s is "2-prover proof systems" and "parallel repetition," which is in some sense related to the amortization of a single instance over many instances. [Raz 1994] changed the landscape by bringing information theory to the forefront in such problems in computer science. [Holenstein] introduced some key methods to prove some important theorems. This is a rather difficult proof to wrap your head around, so reader beware.

Interactive communication with errors (with an adversary choosing the error, or with random errors) is another theme to study.

7.5 Shearer's Lemma and Approximate Independence

Recall Shearer's Lemma: 3d volume is bounded by the square root of the product of 2d volumes. The only possible way to get equality is a "box." The information theory analogue is $H(X, Y, Z)$ bounded by a combination of $H(X, Y)$, $H(Y, Z)$, $H(Z, X)$. The "box" in this case is independent variables (see our proof). What happens when the volume is not quite equal to the bound, but is close? We can get approximate independence in [Ellis et al.].