How to compress interactive communication

Boaz Barak, Mark Braverman, Xi Chen, Anup Rao [BBCR09]

Presentation by Timur Kuzhagaliyev, May 2018

Communication complexity vs. conveyed information

- Amount of information conveyed $\leq$ communication complexity
- Measures of information:
  - **Information cost** (often used in previous work):
    (Amount of) information that an *observer* learns about inputs by observing messages and public randomness.
  - **Information content** (this paper):
    (Amount of) information that *parties* in the protocol learn from observing messages and public randomness, *that they did not already know*.

Definitions from information theory 1

First definition - entropy:

- **Entropy** of a random variable X is defined as

$$H(X) \stackrel{\text{def}}{=} -\sum_{x} \Pr(X = x) log \Pr(X = x).$$

- Conditional entropy is defined as

$$H(X|Y) \stackrel{\text{def}}{=} \sum_{y} \Pr(Y = y) H(X|Y = y).$$

Definitions from information theory 2

Second definition - mutual information:

- **Mutual information** between two random variables A, B, denoted $I(A; B)$, is the quantity

$$I(A;B) \stackrel{\text{def}}{=} H(A) - H(A|B) = H(B) - H(B|A)$$

- Conditional mutual information:

$$I(A;B|C) \stackrel{\text{def}}{=} H(A|C) - H(A|BC) = H(B|C) - H(B|AC)$$

- Chain rule for mutual information, $X^n = X_1, \ldots, X_n$:

$$I(X^n;Y) = \sum_{i=1}^{n} I(X_i; Y \,|\, X_{i-1}, X_{i-2}, \ldots, X_1)$$

Information content

Now we're ready to define information content:

- Given a distribution μ on input X, Y, and protocol π, the **information content** of π is

$$\mathrm{IC}_\mu(\pi) \stackrel{\text{def}}{=} I(X; \pi(X,Y)|Y) + I(Y; \pi(X,Y)|X)$$

  where $\pi(X, Y)$ is the *transcript* of the protocol (concatenation of public randomness and exchanged messages).

Properties of information content

- Each party knows their input, so protocol can only reveal *less* information to them than an independent observer
  $\Rightarrow$ **information content** $\leq$ **information cost**

- When inputs are **independent**, information content is equal information cost

- When inputs are **dependent**, information content can be significantly smaller...

Simple example

Consider the case when:

- μ is a distribution on inputs where we always have $X = Y$

$$\Downarrow$$

- Any communication yields 0 information *content*

- Information *cost* can be arbitrarily large

Limitations in previous work

- Using the notion of **information cost**
  - As we have seen, it isn't always the best measure
- Trying to compress each message separately
  - Inefficient when information content $<< 1$ for every bit of communication (can't afford to transmit even a single bit)
- Methods proposed in our paper eliminate both of these

Compressing communication protocols [1]

- The paper presented two new *protocol compression* methods for protocols with communication complexity C and information content I:

  1. For **non-product** distributions over input, compression down to complexity $\tilde{O}(\sqrt{I \cdot C})$
  2. For **product** distributions over input, compression down to complexity $\tilde{O}(\sqrt{I})$

- It showed that we can transform protocol with **small information content** into a protocol with **small communication complexity** (in expectation).

[1] $f(n) = \tilde{O}(g(n))$ is shorthand for $f(n) = O(g(n) log^k g(n))$ for some k

Compressing communication protocols 2

Need some more definitions:

- Communication complexity of protocol π is denoted $\mathrm{CC}(\pi)$.
- Let D and F be two random variables taking values in a set S. Their **statistical distance** is

$$
\begin{aligned}
|D - F| &\stackrel{\text{def}}{=} \max_{T \subseteq S}(|\Pr(D \in T) - \Pr(F \in T)|) \\
&= \frac{1}{2} \sum_{s \in S} |\Pr(D = s) - Pr(F = s)|
\end{aligned}
$$

Main theorem

Theorem 1.2

There is a universal constant c such that for every distribution μ, every protocol π, every $\epsilon > 0$, there exist functions π_x, π_y and a protocol τ such that:

- $|\pi_x(X, \tau(X, Y)) - \pi(X, Y)| < \epsilon$,
- $\Pr(\pi_x(X, \tau(X, Y)) \neq \pi_y(Y, \tau(X, Y)) < \epsilon$, and

$$\mathrm{CC}(\tau) \leq c\sqrt{\mathrm{CC}(\pi) \cdot \mathrm{IC}(\pi)} \cdot \frac{\log(\mathrm{CC}(\pi)/\epsilon)}{\epsilon}$$

Breakdown of the main theorem

Theorem 1.2

- $|\pi_x(X, \tau(X,Y)) - \pi(X,Y)| < \epsilon$
  - Ensures that transcript of τ specifies a unique leaf that is ϵ-close in statistical distance to the leaf sampled by π
- $\Pr(\pi_x(X, \tau(X,Y)) \neq \pi_y(Y, \tau(X,Y)) < \epsilon$
  - Guarantees with high probability that both players agree on what the sampled leaf was
- Thus the triple τ, π_x, π_y can be used to specify a new protocol that is a compression of π.

Proof of the main theorem 1

- To prove Theorem 1.2, we first consider protocol tree $\mathcal{T}$ for π_r, for every fixing of public randomness r. Let R be the random variable for the public randomness in π.

- **Claim**: $\mathrm{IC}_\mu(\pi) = \mathbb{E}_R[\mathrm{IC}_\mu(\pi_R)]$

Proof of the main theorem 2

Proof:

$$\begin{aligned}
\mathrm{IC}_\mu(\pi) &= I(\pi(X,Y);X|Y) + I(\pi(X,Y);Y|X) \\
&= I(R\pi_R(X,Y);X|Y) + I(R\pi_R(X,Y);Y|X) \\
&= I(R;X|Y) + I(R;Y|X) + I(\pi_R(X,Y);X|YR) \\
&\qquad\qquad\qquad\qquad\qquad\qquad + I(\pi_R(X,Y);Y|XR) \\
&= I(\pi_R(X,Y);X|YR) + I(\pi_R(X,Y);Y|XR) \\
&= \mathop{\mathbb{E}}_R[\mathrm{IC}_\mu(\pi_R)]
\end{aligned}$$

Proof of the main theorem 3

- **Useful trick**: We can describe protocol π_r in a non-standard but equivalent way (see board).

Proof of the main theorem 4

- **Main idea**: Simulate protocol π, trying to avoid communicating by guessing what the other player's sample looks like.
- Players can make many mistakes, but these mistakes can be corrected using the following lemma:

Lemma of Feige et al. [FPRU94]

There is a randomized public coin protocol τ with communication complexity $O(\log(k/\epsilon))$ such that on input two k-bit strings x, y, it outputs the first index $i \in [k]$ such that $x_i \neq y_i$ with probability at least $1 - \epsilon$, if such an i exists.

Proof of the main theorem 5

We define a new protocol $\tau_{\beta,\gamma}$ for some error parameters β, γ, with three phases. Description is for player P_x, but we can just replace x with y to obtain a description for player P_y.

- **Phase 1: Public sampling**
    1. Sample r according to the distribution of public randomness in π.

Proof of the main theorem 6

- **Phase 2: Correlated sampling**:
    1. For every non-leaf node w in the tree, let k_w be uniformly random element of $[0, 1]$ sampled using public randomness.
    2. On input x, player P_x defines the tree $\mathcal{T}_x$ the following way:

       For each node w, player includes the edge to the left child of w if the following holds:

       $$\Pr(\pi_r(X, Y) \text{ reaches left child} \mid \pi_r(X, Y) \text{ reaches } w \text{ and } X = x) > k_w$$

       Otherwise, right child is picked. Note that if P_x owns w, the calculated probability for the left child of w is always "correct".

Proof of the main theorem 7

- **Phase 3: Path finding**:
  1. Each of the players computes the unique path in their trees that leads from the root to a leaf.
  2. Players use lemma from earlier, communicating $O(log(\mathrm{CC}(\pi)/\beta))$ bits to find the first node at which their paths differ (if it exists). The relevant edge is then corrected in favour of the player who owns the node. The player who does not own the node recomputes their path.
  3. Players repeatedly correct their paths $\sqrt{\mathrm{CC}(\pi) \cdot \mathrm{IC}_\mu(\pi)}/\gamma$ times.

Proof of the main theorem 8

- The protocol we defined, $\tau_{\beta,\gamma}$, has the following upper bound for communication complexity by design:

$$\mathrm{CC}(\tau_{\beta,\gamma}) \leq O\left(\sqrt{\mathrm{CC}(\pi) \cdot \mathrm{IC}(\pi)} \cdot \frac{\log(\mathrm{CC}(\pi)/\beta)}{\gamma}\right)$$

- Let $V = V_0, \ldots, V_{\mathrm{CC}(\pi)}$ denote the "right path" in the protocol tree of $\tau_{\beta,\gamma}$, i.e. for every i, V_{i+1} was picked by the owner of V_i. This path has the right distribution, since every child is sampled with exactly the right conditional by the corresponding owner.

Proof of the main theorem 9

- That is, the following claim holds: For every x, y, r, the distribution $V|xyr$ as defined above is the same as the distribution of the sampled transcript in the protocol π.
- This implies:

$$I(X; V|rY) + I(Y; V|rX) = \mathrm{IC}_\mu(\pi_r)$$

Proof of the main theorem 10

We can now show that the expected number of mistakes is small.

- **Claim:**
  $\mathbb{E}[\,\#\text{ of mistakes in simulating } \pi_r \mid r\,] \leq \sqrt{\mathrm{CC}(\pi)\cdot \mathrm{IC}_\mu(\pi_r)}$
- **Proof**: For $i = 1, \ldots, \mathrm{CC}(\pi)$, denote by C_{ir} the indicator random variable for whether or not a mistake occurs at level i, i.e. the edges V_{i-1} are inconsistent in the trees.

Proof (cont.):

- We can bound $\mathbb{E}[C_{ir}]$ for each i: Note that a mistake occurs at the relevant edge when either:
  1. P_x's probability is greater than k_w, while P_y's is smaller than k_w.
  2. P_y's probability is greater than k_w, while P_x's is smaller than k_w.
- Denote the event of the protocol reaching a particular node at level $i-1$ as $v_{<i}$.
- Combining the two cases from before in which a mistake occurs, we get that the probability that a mistake occurs at level i is at most

$$|(V_i|xv_{<i}r) - (V_i|yv_{<i}r)|.$$

Proof of the main theorem 12

Some more definitions:

- Informational **divergence** between two distributions is defined as

$$\mathbb{D}(A \,||\, B) = \sum_x A(x) \log\left(\frac{A(x)}{B(x)}\right)$$

- *Property #1*: $\mathbb{D}(A \,||\, B) \geq |A - B|^2$
- *Property #2*: Let A, B, C be random variables in the same probability space. For every $a \in \mathrm{supp}(A)$ and every $c \in \mathrm{supp}(C)$, let B_a denote "$B|A = a$" and B_{ac} denote "$B|A = a, C = c$". Then

$$I(A; B|C) = \mathop{\mathbb{E}}_{a,c \in_R A,C}[\mathbb{D}(B_{ac} \,||\, B_c)]$$

- Reminders: $\mathbb{E}\left[\sum_i X_i\right] = \sum_i \mathbb{E}[X_i]$ $\qquad$ $\mathbb{E}\left[\sqrt{X}\right] \leq \sqrt{\mathbb{E}[X]}$
- Cauchy Schwartz: $(\mathbb{E}[AB])^2 \leq \mathbb{E}[A]^2\, \mathbb{E}[B]^2$

Proof of the main theorem 13

Proof (cont.):

$$
\begin{aligned}
&\mathbb{E}[C_{ir}] \\
&\le \mathop{\mathbb{E}}_{xyv_{<i} \in_R XYV_{<i}} \left[|(V_i|xv_{<i}r) - (V_i|yv_{<i}r)| \right] \\
&\le \mathop{\mathbb{E}}_{xyv_{<i} \in_R XYV_{<i}} \left[\max\{|(V_i|xyv_{<i}r) - (V_i|yv_{<i}r)|, |(V_i|xyv_{<i}r) - (V_i|xv_{<i}r)|\} \right] \\
&\le \mathop{\mathbb{E}}_{xyv_{<i} \in_R XYV_{<i}} \left[\sqrt{\mathbb{D}(V_i|xyv_{<i}r \,||\, V_i|yv_{<i}r) + \mathbb{D}(V_i|xyv_{<i}r \,||\, V_i|yv_{<i}r)} \right] \\
&\le \sqrt{\mathop{\mathbb{E}}_{xyv_{<i} \in_R XYV_{<i}} \left[\mathbb{D}(V_i|xyv_{<i}r \,||\, V_i|yv_{<i}r) + \mathbb{D}(V_i|xyv_{<i}r \,||\, V_i|yv_{<i}r) \right]} \\
&= \sqrt{I(X; V_i \mid YV_{<i}r) + I(Y; V_i \mid XV_{<i}r)}
\end{aligned}
$$

Proof of the main theorem 14

Proof (cont.): Showing total number of mistakes simulating π_r is small:

$$
\begin{aligned}
\mathbb{E}\left[\sum_{i=1}^{\mathrm{CC}(\pi)} C_{ir}\right] = \sum_{i=1}^{\mathrm{CC}(\pi)} \mathbb{E}[C_{ir}] \qquad &\leq \sum_{i=1}^{\mathrm{CC}(\pi)} \sqrt{(\mathbb{E}[\sqrt{\mathrm{CC}(\pi)} \cdot C_{ir}])^2} \\
&\leq \sqrt{\mathrm{CC}(\pi) \sum_{i=1}^{\mathrm{CC}(\pi)} \mathbb{E}[C_{ir}]^2} \\
&\leq \sqrt{\mathrm{CC}(\pi) \sum_{i=1}^{\mathrm{CC}(\pi)} (I(X; V_i \mid YV_{<i}r) + I(Y; V_i \mid XV_{<i}r))} \\
&= \sqrt{\mathrm{CC}(\pi) \cdot \big(I(X; V^{\mathrm{CC}(\pi)} \mid Yr) + I(Y; V^{\mathrm{CC}(\pi)} \mid Xr)\big)} \\
&= \sqrt{\mathrm{CC}(\pi) \cdot \mathrm{IC}_\mu(\pi_r)}
\end{aligned}
$$

Proof of the main theorem 15

Finally, we can show that the overall total number of mistakes simulating π is small:

$$\begin{aligned}
\mathbb{E}[\ \#\text{ of mistakes in simulating } \pi\ ] &= \mathop{\mathbb{E}}_{R}[\ \#\text{ of mistakes in simulating } \pi_R\ ] \\
&\leq \mathop{\mathbb{E}}_{R}[\sqrt{\mathrm{CC}(\pi)\cdot \mathrm{IC}_\mu(\pi_R)}] \\
&\leq \sqrt{\mathop{\mathbb{E}}_{R}[\mathrm{CC}(\pi)\cdot \mathrm{IC}_\mu(\pi_R)]} \\
&= \sqrt{\mathrm{CC}(\pi)\cdot \mathrm{IC}_\mu(\pi)}
\end{aligned}$$

Proof of the main theorem 16

The protocol fails when both players **do not** finish with the (same) leaf $V_{\mathrm{CC}(\pi)}$, which happens when either:

1. The number of mistakes in the correct path is larger than $\sqrt{\mathrm{CC}(\pi) \cdot \mathrm{IC}_\mu(\pi)}/\gamma$
2. Mistake-correction protocol fails to correct all mistakes

Proof of the main theorem 17

1. The number of mistakes in the correct path is larger than $\sqrt{\mathrm{CC}(\pi) \cdot \mathrm{IC}_\mu(\pi)}/\gamma$
    - Can use Markov's inequality and the result for expected number of errors from before:

$$\begin{aligned}
\Pr[\#\text{ of mistakes} \geq a] &\leq \frac{\mathbb{E}[\ \#\text{ of mistakes }]}{a} \\
&= \frac{\sqrt{\mathrm{CC}(\pi) \cdot \mathrm{IC}_\mu(\pi)}}{\sqrt{\mathrm{CC}(\pi) \cdot \mathrm{IC}_\mu(\pi)}/\gamma} \\
&= \gamma
\end{aligned}$$

2. Mistake-correction protocol fails to correct all mistakes
    - There are $\sqrt{\mathrm{CC}(\pi) \cdot \mathrm{IC}_\mu(\pi)}/\gamma$ correction steps, and correction fails with probability β, giving total error probability of $\beta/\gamma \cdot \sqrt{\mathrm{CC}(\pi) \cdot \mathrm{IC}_\mu(\pi)}$

Proof of the main theorem 18

- By union bound, the probability of failure is bounded by

$$\gamma + \beta/\gamma \cdot \sqrt{\mathrm{CC}(\pi) \cdot \mathrm{IC}_\mu(\pi)}$$

- Set $\beta = \gamma^2/\mathrm{CC}(\pi)$:

$$\begin{aligned}
\gamma + \beta/\gamma \cdot \sqrt{\mathrm{CC}(\pi) \cdot \mathrm{IC}_\mu(\pi)} &= \gamma + \frac{\gamma^2/\mathrm{CC}(\pi)}{\gamma} \cdot \sqrt{\mathrm{CC}(\pi) \cdot \mathrm{IC}_\mu(\pi)} \\
&= \gamma + \gamma \cdot \sqrt{\frac{\mathrm{IC}_\mu(\pi)}{\mathrm{CC}(\pi)}} \\
&\leq 2\gamma
\end{aligned}$$

Proof of the main theorem 19

- Finally, setting $\epsilon = 2\gamma$, we get that the probability of failure is at most ϵ.
- Recall that by design, the protocol has complexity

$$\mathrm{CC}(\tau_{\beta,\gamma}) \leq O\left(\sqrt{\mathrm{CC}(\pi) \cdot \mathrm{IC}(\pi)} \cdot \frac{\log(\mathrm{CC}(\pi)/\beta)}{\gamma}\right)$$

which, with our substitutions, becomes just

$$\mathrm{CC}(\tau_{\beta,\gamma}) \leq O\left(\sqrt{\mathrm{CC}(\pi) \cdot \mathrm{IC}(\pi)} \cdot \frac{\log(\mathrm{CC}(\pi)/\epsilon)}{\epsilon}\right)$$

Wrap up

Thus we have proved the theorem:

Main theorem

There is a universal constant c such that for every distribution μ, every protocol π, every $\epsilon > 0$, there exist functions π_x, π_y and a protocol τ such that:

- $|\pi_x(X, \tau(X, Y)) - \pi(X, Y)| < \epsilon$,
- $\Pr(\pi_x(X, \tau(X, Y)) \neq \pi_y(Y, \tau(X, Y)) < \epsilon$, and

$$\mathrm{CC}(\tau) \leq c\sqrt{\mathrm{CC}(\pi) \cdot \mathrm{IC}(\pi)} \cdot \frac{\log(\mathrm{CC}(\pi)/\epsilon)}{\epsilon}$$

Thanks for listening!

References

[BBCR09] B. Barak, M. Braverman, X. Chen and A. Rao, How to Compress Interactive Communication. *SIAM J. Comput.*, 42(3):1327–1363 2009.

[FPRU94] U. Feige, D. Peleg, P. Raghavan, and E. Upfal. Computing with noisy information. *SIAM Journal on Computing*, 23(5):1001-1018, 1994.