Efficient Semantic Communication & Compatible Beliefs

Madhu Sudan
Microsoft, New England

Based on joint works with:
Oded Goldreich (Weizmann) and Brendan Juba (Harvard)
Part I: Background on Semantic Communication
Semantic Communication: Motivation

- First era of communication: Reliable Wires
  - Essentially done: wires are very reliable; performance can be enhanced (maybe) quantitatively but not qualitatively.
- Can we get endpoints to also be reliable?
  - Modern systems have “smart” endpoints.
  - Smart implies capability.
  - Smart implies diversity.
    - But diversity implies (potential) misunderstanding.
- Semantic Communication [Goldreich,Juba+S ‘10]
  - Detect/Correct Misunderstanding.
Semantic Communication: Model - I

- General Model: Two “smart” interacting agents: User and Server; User wishes to accomplish some Goal, and Server is trying to help User.

- Interacting agents?
  - Agent: State x Inputs → New State x Outputs
Uncertainty about the receiver:
(User doesn’t know server; vice versa).

New Class of Problems
New challenges
Needs more attention!
Semantic Communication: Model - IIIa

- **Goal-oriented communication:**
  - User attempting to reach some goal.
  - How to model this?

- **Classical approaches:**
  - Some function of state of user, or some function of transcript of interaction etc.
  - Fails in “semantic/uncertainty” setting.

- **Our [GJS] approach:** Introduce a (hypothetical) third agent.
Semantic Communication: Model - IIIb

- **Referee**
  - Poses tasks to user.
  - Judges success.

- **Generic Goal specified by** \((R, B, Ŷ, Š)\)
  - \(R\) = Referee (just another agent)
  - \(B\) = Boolean Function determining if the state evolution of the referee reflects successful achievement of goal.
  - \(Ŷ, Š\) = Class of users/servers.
  - (All finitely specified)\(u\)

- Which goals can be achieved universally?
Basic Definitions: Helpfulness, Universality, Sensing

- What makes a server helpful?
  - S is G-helpful, if there exists a user who can achieve goal (efficiently) for every starting state of S.

- Universality:
  - User U is universal if it achieves G with every G-helpful server.

- Sensing?
  - Roughly, Goal G can be sensed if there exists an efficient algorithm that scan use (with their inputs) to see if Referee will accept.
Principal Thesis and Theorem

- Thesis: Every Goal of communication captured in our model (by appropriate choice of \( (R, B, ŭ, ź) \))

- Theorem: Goal is universally achievable if and only if there exists a sensing function (for “one-shot” goals).
Proof + Insights

- Positive results by enumeration.
- Negative? Mostly by definition.

Insights:
- Servers should know how to be “interrupted”. (How else can they function independent of complexity of their own state?)
- Short “interrupt” signal helps.
- Goals should be “sense”ible.
Part II: Beliefs & Compatibility
Motivation

- Why does natural (human) communication differ so much from designed communication?
  - Languages are ambiguous
  - They violate their own grammatical rules
  - They are needlessly redundant at times, and noisily compressed at other times?
- Can we use information theory to explain such phenomena (departures from information theory)?
  - Use fact that natural communication deals with uncertainty about server.
Does Semantic Communication help?

- **Pros:**
  - Does deal with uncertainty about servers.

- **Cons:**
  - Seems quite inefficient (user is enumerating all servers?).
  - Seems to throw away all “knowledge” about server (that might yield efficiency).
  - Is universality really a goal? Is it not at odds with “use of knowledge”
Beliefs in Semantic Communication

- Addition to the model, to include beliefs of user and server.

- Each server and user has associated belief.

- Belief of Server S = \( D_S \) = Distribution on Users

- Belief of User U = \( D_U \) = Distribution on Servers

- Compatibility?
Compatibility of Beliefs

- For user U, Consider the distribution $\mathcal{D}_U$ on users obtained as follows:
  - Sample Server $S'$ from distribution $\mathcal{D}_U$
  - Sample user $U'$ from distribution $\mathcal{D}_{S'}$

- Compatibility of user U with server S:
  - Measured by “proximity” of $\mathcal{D}_U$ with $\mathcal{D}_S$
  - Our choice: U is $\alpha$-compatible with S if
    $$1 - \frac{1}{2} |\mathcal{D}_U - \mathcal{D}_S|_1 \geq \alpha$$
Server Performance?

- What does it mean that server has a belief about users? How is it reflected in server’s actions?

- Performance of server S (roughly) for goal G
  \[ \text{Performance of server } S \text{ (roughly) for goal } G = \text{Perf}_G(S) = \text{Expected time that user } U \text{ chosen from distribution } D_S \text{ takes to achieve goal } G. \]

- Well-designed server should be “broad-minded” i.e., efficient against a broad distribution of users.
Universality of Users under Beliefs

- Universal User U now has beliefs on servers.

- Can expect user to do well not only on servers in the support of its beliefs, but a potentially broader set: namely servers with compatible beliefs.

- Theorem [Juba, S ’11]: \( \forall U, \exists U' \) a universal user U’ with same beliefs as U, whose time to achieve goal G with server S is \( \frac{O(1)}{\alpha(U,S)} \text{Perf}_G(S) \), provided the goal allows universal users.
Consequences/Conclusions

- Universality of communication is not at odds with efficiency.

- Efficiency comes with compatibility of communicating players.

- Universality takes care of possibility of misunderstanding, at an affordable price.
Thank you