Distributed Functional Compression through Graph Coloring

Vishal Doshi, Devavrat Shah, Muriel Medard

Massachusetts Institute of Technology
Distributed Functional Compression

- Nodes (sensors, cameras) in a network have pieces of information
  - Different nodes have different use of the information in network
  - i.e. they wish to compute different functions of the information

- Question: what is the minimal transmission rates between nodes
  - So as to allow different nodes to compute their functions
  - And, nodes have to do this in a distributed manner
Applications

• Distributed function compression
  – Marriage of information theory, networks and computation
  – Grand challenge for modern information theory

• Algorithmic applications
  – Surveillance cameras gather a lot of information
  – But only small amount of information is useful
  – Distributed compression can lead to an efficient network

• Policy guidelines
  – Privacy preservation of sensitive information, e.g. Census DB
  – Information theoretic results provide bounds on “information transmission” so as to preserve privacy of people
  – Or, Privacy preservation of people while using cameras for surveillance in public places
First Step: Functional Side Information

- At what rate must $X$ be encoded such that the computation of $f(X, Y)$ is possible at the receiver?
  - Here, $f$ can be any deterministic function

- More importantly,
  - Can we come up with “implementable” coding schemes to achieve (close to) the optimal rate?
Past Work

• The minimal rate was characterized by Orlitsky-Roche (2001):

\[ H_{\min}^{Y \mid X; W \mid X, Y} \equiv ( \mid ) \]

where G is a graph defined by the function f(X,Y) and the distribution p(X,Y) called the characteristic graph.

• \( H_G(X \mid Y) < H(X \mid Y) \) and the disparity is determined by the complexity of the function
  – Thus, functional compression definitely provides gains

• However, the above scheme is somewhat complicated
  – We provide coding scheme using notion of graph coloring
  – It can achieve the optimal rate
  – Leads to design of good simple-to-implement heuristics
Our Coding Scheme

- The following scheme achieves the optimal rate:

1. Construct $G^n$ and color it (gains even without min entropy colorings)
2. Color each source symbol
3. Do Slepian-Wolf style coding to recover $c(X^n)$ and $Y^n$ at the receiver
4. With high probability, $c(X^n)$ and $Y^n$ uniquely determine $f(X^n, Y^n)$

Our scheme is built on existing distributed source coding
- Provides separation between “functional coding” and “data
Step II: Distributed Functional Compression

- Next we consider the distributed source coding set up

- Again, we wish to determine minimal transmit rates for $X$ and $Y$ such that $f(X,Y)$ can be recovered at the receiver?
  - The sources cannot communicate with each other

- Can we extend our result for Step I to this more general problem?
  - Answer: Yes, but with some conditions
Distributed Functional Compression

- Condition: the joint distribution of sources is such that
  - $p(x_1, y_1) > 0$, $p(x_2, y_2) > 0$ implies $p(x_1, y_2) > 0$
  - This is a reasonable assumption for a wide array of applications

- Under this condition, our coding scheme is optimal

- The rate region is the closure of the set of all $(R_X, R_Y)$ such that for some $n$ and colorings $c_X$ and $c_Y$ of $G^n$ and $H^n$, where $H^n$ is the characteristic graph of $Y$ with respect to $X$, the following holds:

  \[ nR_X \geq H(c_X(X)|c_Y(Y)) \]
  \[ nR_Y \geq H(c_Y(Y)|c_X(X)) \]
  \[ n(R_X + R_Y) \geq H(c_X(X), c_Y(Y)) \]
Our Coding Scheme

• The following scheme achieves the optimal rate:

1. Construct graphs $G^n$ and $H^n$ and color them

2. Color each source symbol

3. Do Slepian-Wolf style coding to recover $c_X(X^n)$ and $c_Y(Y^n)$ at receiver

4. With high probability, $c_X(X^n)$ and $c_Y(Y^n)$ uniquely determine $f(X^n, Y^n)$
Rate Region for Distributed Coding

- The inner region is the classical Slepian-Wolf rate region.
- We can achieve rates out to the blue curve using the function information.
- We have single letter characterizations for three points on the region including the minimal joint rate point.
Summary

• We are developing novel coding schemes that utilize graph coloring as a means to achieve optimal rates

• Our results extends to multiple sources and one receiver

• Our approach leads to layered architecture
  – Add “functional layer” on existing distributed compression schemes

• There is a lot of literature on Approximation algorithms for graph coloring
  – This allows for heuristics based on these algorithms

• And, it furthers the current project in the correct direction!