Perspectives on Information Bottleneck Problems

Shlomo Shamai

Technion—Israel Institute of Technology
sshlomo@ee.technion.ac.il

Joint work with: Abdellatif Zaidi (Université Paris-Est, Paris)

Information Theory and Applications
February 2019 San Diego, CA, USA

Outline

* **Information Bottleneck:**
* **Connections:**
  - Remote Source Coding.
  - Common Reconstruction.
  - Information Combining.
  - Wyner-Ahlswede-Korner Problem.
  - Efficiency of Investment Information.
* **Distributed Information Bottleneck:**
  - CEO Source Coding Problem under Log-Loss.
  - Oblivious Relay Processing (7,8).
  - Distributed Information Bottleneck for Learning.
* **A Short Outlook**

Information Bottleneck

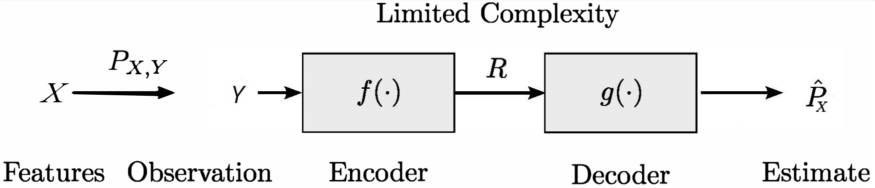
$$X \longrightarrow Y \longrightarrow U$$

- Efficiency of a given representation $U = f(Y)$ measured by the pair

  Rate (or *Complexity*): $I(U;Y)$ and Information (or *Relevance*): $I(U;X)$

- Information $I(X;U)$ can be achieved by OBLIVIOUS coding Y while with the logarithmic distortion with respect to X
- Single letter-wise, U is not necessarily a deterministic function of Y
- The non-oblivious bottleneck problem is immediate as the $\min(I(X;Y), R)$ is achievable by having the relay decoding the message transmitted by X
- The bottleneck problem connects to many timely aspects, such as 'deep learning' *[Tishby-Zaslavsky, ITW'15]*.

Digression: Learning via the Information Bottleneck Method



- Preserving all the information about X that is contained in Y, i.e., $I(X;Y)$, requires high *complexity* (in terms of *minimum description coding length*).
  - Other measures of complexity may be (Vapnik-Chervonenkis) VC-dimension, covering numbers, ..
- Efficiency of a given representation $U = f(Y)$ measured by the pair

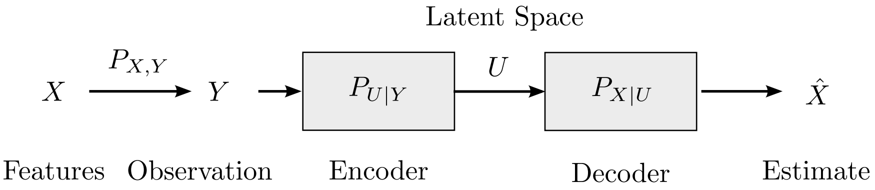
$$\text{Complexity:} \quad I(U;Y) \qquad \text{and} \qquad \text{Relevance:} \quad I(U;X)$$

- Example:

$$\max_{p(u|x)} I(U;X) \quad \text{s.t.} \quad I(U;Y) \leq R, \quad \text{for} \quad 0 \leq R \leq H(Y)$$

$$\min_{p(u|x)} I(U;Y) \quad \text{s.t.} \quad I(U;X) \geq \Delta, \quad \text{for} \quad 0 \leq \Delta \leq I(X;Y)$$

Basically, a Remote Source Coding Problem !



- Reconstruction at decoder is under log-loss measure,

$$R(\Delta) = \min_{p(u|y)} I(U;Y)$$

where the minimization is over all conditional pmfs $p(u|y)$ such that

$$\mathbb{E}[\ell_{\mathsf{log}}(X,U)] \leq H(X) - H(X|U) = H(X) - \Delta$$

- R. L. Dobrushin and B. S. Tsybakov, *"Information transmission with additional noise"*, IRE Tran. Info. Theory, Vol. IT-8, pp. 293-304, 1962.

- H. Witsenhausen, A. Wyner, *"A conditional entropy bound for a pair of discrete random variables"*, IEEE Trans. on Info. Theory, Vol. 21, pp. 493-501, 1975.

- Solution also coined as the Information Bottleneck Method *[Tishby'99]*

$$L_{\mathrm{IB}}(\beta, P_{X,Y}) = \min_{p(u|y)} I(Y;U) - \beta I(X;U)$$

Other Connections

- **The Efficiency of Investment Information**
  - X - Stock Market Data.

    Y - Correlated Information about X.

    $\Delta(C)$ the maximum increase in growth rate when Y is described to the investor at rate C (a logarithmic distortion that relates to the Wyner-Ahlswede-Korner Problem).

  - Solution of the bottleneck for: (X, Y) are binary and (X, Y) Gaussian (horse race examples).

  - E. Erkip and T. M. Cover, "The Efficiency of Investment Information", IEEE Trans. on Info. Theory, Vol. 44, May 1975.

Other Connections (Cont.)

- **Common Reconstruction**. Because $X -\!\!\circ\!\!- Y -\!\!\circ\!\!- U$, we have

$$\begin{aligned} I(U;X) &= I(U;Y) - I(U;Y|X) \\ &\leq R - I(U;Y|X) \end{aligned}$$

- Y. Steinberg, *"Coding and common reconstruction"*, IEEE Trans. on Info. Theory, vol. 55, no. 11, pp. 4995–5010, Nov. 2009 (X – side information is not used for the 'source' Y common reconstruction).

∗ Heegard-Berger Problem with Common Reconstruction: Y-source, to be commonly reconstructed (with logarithmic distortion), with and without side information (X), as to maximize $I(U;X)$.

- M. Benammar, A. Zaidi, "Rate-Distortion of a Heegard-Berger Problem with Common Reconstruction Constraint," IZS, March 2–4, 2016.

Other Connections (Cont.)

- **Information Combining**

  $I(Y;U,X) = I(U;Y) + I(X;Y) - I(U;X)$ (since $U \multimap Y \multimap X$)

  Since $I(X;Y)$ is given and $I(Y;U) = R$, maximizing $I(U;X)$ is equivalent to minimizing $I(Y;U,X)$.

- I. Sutskover, S. Shamai and J. Ziv, "Extremes of Information Combining", IEEE Trans. Inform. Theory, vol. 51, no. 4, pp. 1313–1325, April 2005.
- I. Land and J. Huber, *"Information combining,"* Foundations and trends in Commun. and Inform. Theory, vol. 3, pp. 227–330, Nov. 2006.

Other Connections (Cont.)

- **Elegant Proofs of Classical Bottleneck Results**

  - X, Y binary symmetric connected through a Binary Symmetric Channel (error probability e): U-Y, also a BSC, $I(U;X) = \{1 - h(e^*v)\}$ where $e^*v = e(1-v) + v(1-e), R = 1 - h(v)$.

  Directly extends to $X - Y$ symmetric, where Y is symmetric binary (one bit output quantization).

  - X Gaussian, and $Y = \sqrt{(\mathrm{snr})})X + N$ (N standard Gaussian). Elegant proof via I-MMSE [Guo-Shamai-Verdu, FnT'13].

$$I(U;X) = \frac{1}{2}\log(1+\mathrm{snr}) - \frac{1}{2}\log\Big(1 + \mathrm{snr}\exp(-2R)\Big)$$

Other Connections (Cont.)

- **Wyner-Ahlswede-Körner Problem**

  If X and Y are encoded at rates R_X and R_Y, respectively. For given $R_Y = R$, the minimum rate R_X that is needed to recover X losslessly is

$$R_X^\star(R) = \min_{p(u|y)\,:\,I(U;Y)\,\leq\,R} H(X|U)$$

  So, we get

$$\max_{p(u|y)\,:\,I(U;Y)\leq R} I(U;X) = H(X) - R_X^\star(R)$$

- R. F. Ahlswede and J. Korner, *"Source coding with side information and a converse for degraded broadcast channels"*, IEEE Trans. on Info. Theory, Vol. 21, pp. 629-637, 1975.

- A. D. Wyner, *"On source coding with side information at the decoder"*, IEEE Trans. on Info. Theory, Vol. 21, pp. 294-300, 1975.

Vector Gaussian Information Bottleneck

- $(\mathbf{X}, \mathbf{Y})$ jointly Gaussian, $\mathbf{X} \in \mathbb{R}^N$ and $\mathbf{Y} \in \mathbb{R}^M$
- Optimal encoding $P_{U|Y}$ is a noisy linear projection to a subspace whose dimensionality is determined by the bottleneck Lagrangian multiplier β *[Chechik et al. '05]*

$$\mathbf{U} = \mathbf{A}\mathbf{Y} + \mathbf{Z}, \qquad \mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$

where

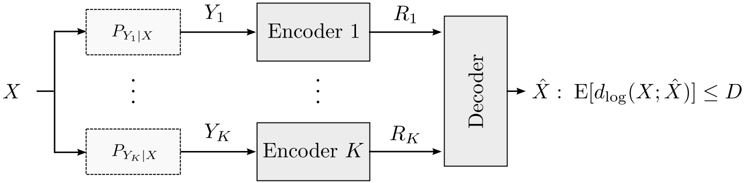
$$\mathbf{A} = \begin{cases} [\mathbf{0}^T; \ldots; \mathbf{0}^T], & \text{if } 0 \leq \beta \leq \beta_1^c \\ [\alpha_1 \mathbf{v}_1^T; \mathbf{0}^T; \ldots; \mathbf{0}^T], & \text{if } \beta_1^c \leq \beta \leq \beta_2^c \\ [\alpha_1 \mathbf{v}_1^T; \alpha_2 \mathbf{v}_2^T; \mathbf{0}^T; \ldots; \mathbf{0}^T], & \text{if } \beta_2^c \leq \beta \leq \beta_3^c \\ \vdots & \end{cases}$$

and $\{\mathbf{v}_1^T, \ldots, \mathbf{v}_N^T\}$ are the left eigenvectors of $\mathbf{\Sigma}_{y|x}\mathbf{\Sigma}_y^{-1}$, sorted by their ascending eigenvalues $\{\lambda_1, \ldots, \lambda_N\}$; $\beta_i^c = 1/(1-\lambda_i)$ are critical β values; $r_i = \mathbf{v}_i^T \mathbf{\Sigma}_y \mathbf{v}_i$ and

$$\alpha_i = \sqrt{\frac{\beta(1-\lambda_i) - 1}{\lambda_i r_i}}$$

Rate-Information Trade-off Gaussian Vector Channel [Winkelbauer-Matz, ISIT'14].

CEO Source Coding Problem under Log-Loss



- CEO source coding problem under log-loss distortion:

$$d_{\log}(x,\hat{x}) := \log\left(\frac{1}{\hat{x}(x)}\right)$$

where $\hat{x} \in \mathcal{P}(\mathcal{X})$ is a probability distribution on $\mathcal{X}$.

- Characterization of rate-distortion region in *[Courtade-Weissman'14]*
  - Key step: log-loss admits a lower bound in the form of conditional entropy of the source conditioned on the compression indices:

$$nD \geq \mathrm{E}[d_{\log}(X^n;\hat{X}^n)] \geq H(X^n|J_{\mathcal{K}}) = H(X^n) - I(X^n;J_{\mathcal{K}})$$

CEO Source Coding Problem under Log-Loss (Cont.)

- Converse of Theorem 1 for Oblivious CRAN leverages on this relation applied to multiple channel inputs, which can be designed.
  Multiple description CEO problem-logloss distortion [Pichler-Piantanida-Matz, ISIT'17].

- Vector Gaussian CEO Problem Under Logarithmic Loss and Applications [Ugur-Aguerri-Zaidi, arxiv:1811.03933]: Accounts also for Gaussian side information about the source at the decoder.

- Full characterization
  (not the case for MMSE Distortion, [Ekrem-Ulukos, IT0214]).

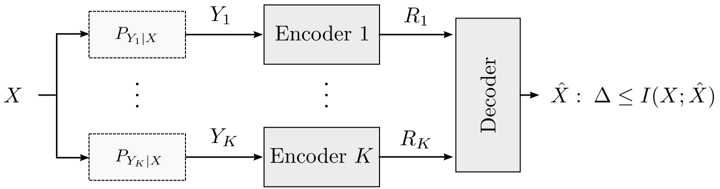
- Implications [Ugur-Aguerri-Zaidi, arxiv:1811.03933] Solutions of:

- Vector Gaussian distributed hypothesis testing against conditional independence [Rahman-Wagner, IT2012].

- A quadratic vector Gaussian CEO problem with determinant constraint.

- Vector Gaussian distributed Information Bottleneck Problem.

Distributed Information Bottleneck



- Information Bottleneck introduced by *[Tishby'99]* and *[Witsenhausen'80]* "Indirect Rate Distortion Problems", IT–26, no. 5, pp. 518–521, Sept. 1980.
- It is a CEO source-coding problem under log-loss!

Theorem (Distributed Information Bottleneck [Estella-Zaidi, IZS'18])

The D-IB region is the set of all tuples $(\Delta, R_1, \ldots, R_K)$ *which satisfy*

$$\Delta \leq \sum_{k \in \mathcal{S}} [R_k - I(Y_k; U_k | X, Q)] + I(X; U_{\mathcal{S}^c} | Q), \qquad \textit{for all } \mathcal{S} \subseteq \mathcal{K}$$

for some joint pmf $p(q)p(x) \prod_{k=1}^{K} p(y_k|x) \prod_{k=1}^{K} p(u_k|y_k, q)$.

Vector Gaussian Distributed Information Bottleneck

- $(\mathbf{Y}_1, \cdots, \mathbf{Y}_K, \mathbf{X})$ jointly Gaussian, $\mathbf{Y}_k \in \mathbb{R}^N$ and $\mathbf{X} \in \mathbb{R}^M$,

$$\mathbf{Y}_k = \mathbf{H}_k \mathbf{X} + \mathbf{N}_k, \qquad \mathbf{N}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{\mathbf{n}_k})$$

- Optimal encoding $P^*_{U_k|Y_k}$ is Gaussian and $Q = \emptyset$ *[Estella-Zaidi'17]*

Theorem ([Estella-Zaidi, IZS'18], [Ugur-Aguerri-Zaidi, arxiv:1811.03933])

If $(\mathbf{X}, \mathbf{Y}_1, \ldots, \mathbf{Y}_K)$ are jointly Gaussian, the D-IB region is given by the set of all tuples $(\Delta, R_1, \ldots, R_L)$ satisfying that for all $\mathcal{S} \subseteq \mathcal{K}$

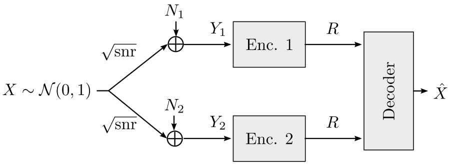
$$\Delta \leq \sum_{k \in \mathcal{S}} \left[R_k + \log |\mathbf{I} - \mathbf{B}_k| \right] + \log \left| \sum_{k \in \mathcal{S}^c} \bar{\mathbf{H}}_k^H \mathbf{B}_k \bar{\mathbf{H}}_k + \mathbf{I} \right|$$

for some $\mathbf{0} \preceq \mathbf{B}_k \preceq \mathbf{I}$, where $\bar{\mathbf{H}}_k = \mathbf{\Sigma}_{\mathbf{n}_k}^{-1/2} \mathbf{H}_k \mathbf{\Sigma}_{\mathbf{x}}^{1/2}$, and achievable with

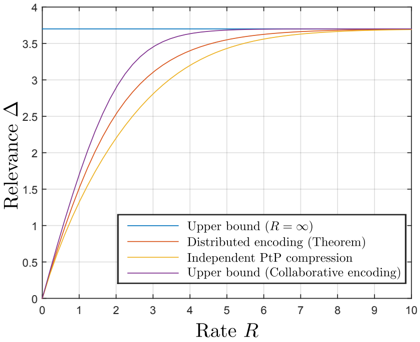
$$p^*(\mathbf{u}_k | \mathbf{y}_k, q) = \mathcal{CN}(\mathbf{y}_k, \Sigma_{\mathbf{n}_k}^{1/2} (\mathbf{B}_k - \mathbf{I}) \mathbf{\Sigma}_{\mathbf{n}_k}^{1/2})$$

- Reminiscent of the sum-capacity in Gaussian Oblivious CRAN with Constant Gaussian Input constraint.

Example



$Y_k = \sqrt{\mathrm{snr}}X + N_k$, with $N_k \sim \mathcal{N}(0,1)$, $k = 1, 2$



- Optimal information (relevance):

$$\Delta^*(R, \mathrm{snr}) = \log_2\left(1 + 2\,\mathrm{snr}\,2^{-2R}\left(2^{2R} + \mathrm{snr} - \sqrt{\mathrm{snr}^2 + (1 + 2\,\mathrm{snr})\,2^{2R}}\right)\right)$$

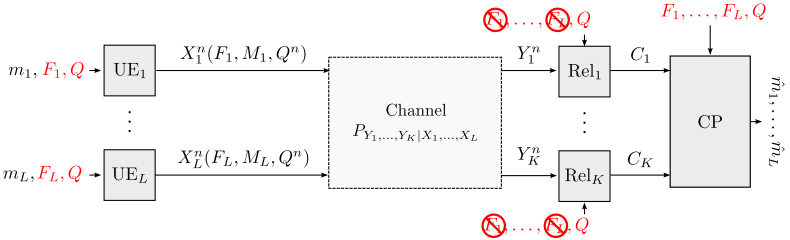
- Collaborative encoding upper bound: (Y_1, Y_2) encoded at rate $2R$

$$\Delta_{\mathrm{ub}}(R, \mathrm{sr}) = \log_2(1 + 2\,\mathrm{snr}) - \log_2(1 + 2\,\mathrm{snr}\,2^{-2R})$$

- Lower bound: Y_1 and Y_2 independently encoded

$$\Delta_i(R, \mathrm{snr}) = \log_2(1 + 2\,\mathrm{snr} - \mathrm{snr}\,2^{-R}) - \log_2(1 + \mathrm{snr}\,2^{-R})$$

Oblivious Relay Processing



- Resource-sharing random variable Q^n available at all terminals *[Simeone et al'11]*.
- Q^n way easier to share, (e.g., on/off activity).
- Memoryless Channel: $P_{Y_1,\ldots,Y_K|X_1,\ldots,X_1}$
- User $l \in \{1, \ldots, L\}$: $\phi_l^n : [1, |\mathcal{X}_l|^{n2^{nR_l}}] \times [1, 2^{nR_l}] \times \mathcal{Q}^n \to \mathcal{X}_l^n$
- Relay $k \in \{1, \ldots, K\}$: $g_k^n : \mathcal{Y}_k{}^n \times \mathcal{Q}^n \to [1, 2^{nC_k}]$
- Decoder:

$$\psi^n : [1, |\mathcal{X}_1|^{n2^{nR_1}}] \times \cdots \times [1, 2^{nC_K}] \times \mathcal{Q}^n \to [1, 2^{nR_1}] \times \ldots \times [1, 2^{nR_L}]$$

Capacity Region of a Class of CRAN Channels

Theorem (Aguerri-Zaidi-Caire-Shamai '17)

For the class of discrete memoryless channels satisfying

$$Y_k -\!\!\circ\!\!- X_{\mathcal{L}} -\!\!\circ\!\!- Y_{\mathcal{K}\setminus k}$$

with oblivious relay processing and enabled resource-sharing, a rate tuple $(R_1, \dots, R_L)$ *is achievable if and only if for all* $\mathcal{T} \subseteq \mathcal{L}$ *and for all* $\mathcal{S} \subseteq \mathcal{K}$*,*

$$\sum_{t\in\mathcal{T}} R_t \leq \sum_{s\in\mathcal{S}} [C_s - I(Y_s; U_s | X_{\mathcal{L}}, Q)] + I(X_{\mathcal{T}}; U_{\mathcal{S}^c} | X_{\mathcal{T}^c}, Q),$$

for some joint measure of the form

$$P_Q \prod_{l=1}^{L} P_{X_l|Q} \prod_{k=1}^{K} P_{Y_k|X_{\mathcal{L}}} \prod_{k=1}^{K} P_{U_k|Y_k,Q},$$

with the cardinality of Q *bounded as* $|\mathcal{Q}| \leq K + 2$*.*

The Distributed Information Bottleneck for Learning

- For simplicity, we look at the D-IB under sum-rate *[Estella-Zaidi'18]*

$$P^*_{U_k|Y_k} = \arg \min_{P_{U_k|Y_k}} I(X; U_{\mathcal{K}}) + \beta \sum_{k=1}^{K} [I(Y_k; U_k) - I(X; U_k)]$$

- The optimal encoders-decoder of the D-IB under sum-rate constraint satisfy the following self consistent equations,

$$\begin{aligned} p(u_k|y_k) &= \frac{p(u_k)}{Z(\beta, u_k)} \exp\left(-\psi_s(u_k, y_k)\right), \\ p(x|u_k) &= \sum_{y_k \in \mathcal{Y}_k} p(y_k|u_k) p(x|y_k) \\ p(x|u_1, \ldots, u_K) &= \sum_{y_{\mathcal{K}} \in \mathcal{Y}_{\mathcal{K}}} p(y_{\mathcal{K}}) p(u_{\mathcal{K}}|y_{\mathcal{K}}) p(x|y_{\mathcal{K}}) / p(u_{\mathcal{K}}) \end{aligned}$$

where

$$\psi_s(u_k, y_k) := D_{\mathrm{KL}}(P_{X|y_k} || Q_{X|u_k}) + \frac{1}{s} \mathrm{E}_{U_{\mathcal{K}\setminus k}|y_k} [D_{\mathrm{KL}}(P_{X|U_{\mathcal{K}\setminus k}, y_k} || Q_{X|U_{\mathcal{K}\setminus k}, u_k}))].$$

- Alternating iterations of these equations converge to a a solution for any initial $p(u_k|x_k)$, similarly to a Blahut-Arimoto algorithm.

D-IB for Vector Gaussian Sources: Iterative Optimization

- $(\mathbf{Y}_1, \cdots, \mathbf{Y}_K, \mathbf{X})$ jointly Gaussian, $\mathbf{Y}_k \in \mathbb{R}^N$ and $\mathbf{X} \in \mathbb{R}^M$,

$$\mathbf{Y}_k = \mathbf{H}_k \mathbf{X} + \mathbf{N}_k, \qquad \mathbf{N}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$

- Optimal encoding $P^*_{U_k|Y_k}$ is Gaussian *[Estella-Zaidi'17]* and given by

$$\mathbf{U}_k = \mathbf{A}_k \mathbf{Y}_k + \mathbf{Z}_k, \qquad \mathbf{Z}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{z,k})$$

- For this class of distributions, the updates in the Blahut-Arimoto type algorithm simplify to:

$$\begin{aligned}
\mathbf{\Sigma}_{\mathbf{z}_k^{t+1}} &= \left(\left(1 + \frac{1}{\beta}\right) \mathbf{\Sigma}^{-1}_{\mathbf{u}_k^t|\mathbf{x}} - \frac{1}{s} \mathbf{\Sigma}^{-1}_{\mathbf{u}_k^t|\mathbf{u}^t_{\mathcal{K}\setminus k}} \right)^{-1}, \\
\mathbf{A}_k^{t+1} &= \mathbf{\Sigma}^{-1}_{\mathbf{z}_k^{t+1}} \left(\left(1 + \frac{1}{\beta}\right) \mathbf{\Sigma}^{-1}_{\mathbf{u}_k^t|\mathbf{x}} \mathbf{A}_k^t (\mathbf{I} - \mathbf{\Sigma}_{\mathbf{y}_k|\mathbf{x}} \mathbf{\Sigma}^{-1}_{\mathbf{y}_k}) \right. \\
&\quad \left. - \frac{1}{\beta} \mathbf{\Sigma}^{-1}_{\mathbf{u}_k^t|\mathbf{u}^t_{\mathcal{K}\setminus k}} \mathbf{A}_k^t (\mathbf{I} - \mathbf{\Sigma}_{\mathbf{y}_k|\mathbf{u}^t_{\mathcal{K}\setminus k}} \mathbf{\Sigma}^{-1}_{\mathbf{y}_k}) \right).
\end{aligned}$$

Short Outlook

- Optimal input distributions for the input power constrained Gaussian bottleneck model.
  - Discrete signaling is already known to sometimes outperform Gaussian signaling for single-user Gaussian CRAN *[Sanderovich et al. '08]*.
  - It is conjectured that the optimal input distribution is discrete.
  - Improved upper bounds (over cut-set) for non-oblivious relay based schemes, to better evaluate the cost of oblivious processing (á la: Vu-Barnes-Ozgur, arXiv:1701.02043 Gaussian primitive relay).
- Connections between classical bottleneck problems and Common Information *[Wyner'75]*, *[Gacs-Korner '73]*.
  - Lossy common information *[Viswanatha-Akyol-Rose, IT2014]*.
  - Network source-coding *[Gray-Wyner'74]*.
  - Information Decomposition, Common Information and Bottleneck *[Banerjee, arXiv: 1503.00709]*.

Short Outlook cont.'

- Bounds on general information bottleneck problems [Painsky-Tishby, `arXiv:1711.02421`], [Eswaran-Gastpar, `arXiv:1805.06515`].
- A variety of related C-RAN & Distributed bottleneck problems:
  - Impact of block length n [C may not scale linearly with $n \Rightarrow$ Courtade conjecture $(C = 1)$] relates to [Courtade-Kumar, IT'14], [Yang-Wesel, arXiv:1807.11289, July'18], The $C = n - 1$ relates to [Huleihel-Ordentlich, arXiv:1701.03119v2, May'17].
  - Bandlimited time-continuous models [Homri-Peleg-Shamai, TCOM, Nov.'18].
  - Multi-layer Information Bottleneck Problem (Yang-Piantanida-Gündüz, arXiv:1711.05102).
  - Distributed Information-Theoretic Clustering (Pichler-Piantanida-Matz, arXiv:1602.04605, Dictator Functions, arXiv:1604.02109).
  - For: $V - X - Y - U$, find:

$$\max I(U;V) \text{ subjected to: } I(V;X) \leq C_1, I(U;Y) \leq C_2 \; .$$

Short Outlook cont.'

- Entropy constraint bottleneck:

  $X - Y - U$
  $\max I(X;U)$ under the constraint $H(U) \leq C$ practical applications:
  LZ distortionless compression.

  $\Rightarrow U = f(y)$ a deterministic function [Homri-Peleg-Shamai, TCOM, Nov.'18]

  – With resource sharing $Q \Rightarrow \max I(X;U|Q)$ subjected to: $H(U|Q) \leq C$.

  - The deterministic bottleneck: advantages in complexity as compared to a classical bottleneck: [Strouse-Schwab, arXiv:1604.00268].

Short Outlook cont.'

- Privacy Funnel, dual of bottleneck: $X - Y - U$, minimize: $I(X;U)$, under the constraint: $I(Y;U) = C$. *[Calmon-Makhdoumi-Medard-Varia-Christiansen-Duffy IT2017]*.
  - Direct connection to Information combining, maximize: $I(Y;U,X) = I(X;Y) + I(U;Y) - I(U;X)$, under the constraint: $I(U;Y) = C$.
  - Example: (X,Y) binary symmetric connected via a BSC, $X - Y$. The channel $Y - U$ is an Erasure Channel.
  - Example: For the Gaussian model: $Y = \sqrt{(\mathrm{snr})}\,X + N$, where (X,N) are unit norm independent Gaussians: Take U to be a deterministic function of Y, say describes the m last digits of a b long $(b \to \infty)$ binary description of Y, such that $I(U;Y) = H(U) = C$ (m is C dependent). Evidently $I(U;X) \to 0$, as $I(Y;U,X) \to C + I(X;Y)$.

References

I. E. Aguerri and A. Zaidi, "Distributed Information Bottleneck Method for Discrete and Gaussian Sources," 2018 International Zurich Seminar on Information and Communication, Zurich, 21–23, February 2018.

R. F. Ahlswede and J. Körner, "Source coding with side information and a converse for degraded broadcast channels", IEEE Trans. on Info. Theory, Vol. 21, pp. 629–637, 1975.

P.K. Banerjee, "Some New Insights into Information Decomposition in Complex Systems Based on Common Information", arXiv:1503.00709, 2015.

M. Benammar and A. Zaidi, "Rate-Distortion of a Heegard-Berger Problem with Common Reconstruction Constraint," Zürich Seminar on Communications and Information, Zürich, Switzerland, March 2–4, 2016.

F. P. Calmon, A. Makhdoumi, M. Médard, M. Varia, M. Christiansen, and K. R. Duffy, "Principal inertia components and applications", vol. 63, no. 9, pp. 5011–5038, 2017.

G. Chechik, A. Globerson, N. Tishby and Y. Weiss, "Information Bottleneck for Gaussian Variables", Journal of Machine Learning Research 6 (2005), pp. 165–188.

G. Chechik, A. Globerson, N. Tishby, and Y. Weiss, "Information bottleneck for Gaussian variables," Journal of Machine Learning Research, vol. 6, pp. 165–188, Feb. 2005.

T.A. Courtade and G.R. Kumar, "Which Boolean Functions Maximize Mutual Information on Noisy Inputs?", IEEE Trans. on Information Theory, vol. 60, pp. 4515–4525, Aug. 2014.

T. A. Courtade and T. Weissman, "Multiterminal source coding under logarithmic loss", IEEE Trans. Inf. Theory, vol. 60, pp. 740–761, Jan. 2014.

R. Dobrushin and B. Tsybakov, "Information transmission with additional noise," IRE Transactions on Information Theory, vol. 8, no. 5, pp. 293–304, September 1962.

E. Ekrem and S. Ulukus, "An Outer Bound for the Vector Gaussian CEO Problem", IEEE Trans. Inform. Theory, vol. 60, no. 11, pp. 6870–6887, November 2014.

E. Erkip and T. M. Cover, "The Efficiency of Investment Information," IEEE Transactions Information Theory, vol. 44, pp. 1026–1040, May 1998.

References (cont.)

K. Eswaran and M. Gastpar, "Remote Source Coding under Gaussian Noise: Dueling Roles of Power and Entropy Power", arXiv:1805.06515.

P. Gacs and J. Körner, "Common information is much less than mutual information," Problems of Control and Information Theory, vol. 2, pp. 149–162, 1973.

P. Gilad-Bachraf, A. Navot, and N. Tishby, "An information theoretic tradeoff between complexity and accuracy," in Proc. COLT, 2003, pp. 595–609. R. Gray and A. Wyner, "Source coding for a simple network", Bell systems Technical Journal, vol. 53, no. 9, pp. 1681–1721, 1974.

D. Guo, S. Shamai, and S. Verdú, "The interplay between information and estimation measures," Foundations and Trends in Signal Processing, vol. 6, pp. 243–429, 2013.

R. M. Hecht and N. Tishby, "Extraction of relevant speech features using the information bottleneck method," in Proc. of InterSpeech, 2005, pp. 353–356.

A. Homri, M. Peleg and S. Shamai (Shitz), "Oblivious Fronthaul-Constrained Relay for a Gaussian Channel", IEEE Trans. on Communications, vol. 66, no. 11, November 2018, pp. 5112–5123.

W. Huleihel and O. Ordentlich, "How to Quantize n Outputs of a Binary Symmetric Channel to n–1 Bits?", IEEE International Symposium on Information Theory (ISIT2017), Aachen, Germany, June 25–30, 2017.

I. Land and J. Huber, "Information combining", Foundations and trends in Commun. and Inform. Theory, vol. 3, pp. 227–330, Nov. 2006.

A. Painsky and N. Tishby, "Gaussian Lower Bound for the Information Bottleneck Limit", arXiv:1711.02421.

S.-H. Park, O. Simeone, O. Sahin and S. Shamai (Shitz), "Fronthaul compression for cloud radio access networks", IEEE Sig. Proc. Mag., Special Issue on Signal Processing for the 5G Revolution, vol. 31, pp. 69–79, Nov. 2014.

G. Pichler and G. Koliander, "Information Bottleneck on General Alphabets", arXiv:1801.01050.

References (cont.)

G. Pichler, P. Piantanida and G. Matz, “A multiple description CEO problem with log-loss distortion”, Proc. IEEE Int. Symp. Inform. Theory, Aachen, Germany, June 25–30, 2017.

G. Pichler, P. Piantanida and G. Matz, “Two Dictator Functions Maximize Mutual Information”, arXiv:1604.02109.

G. Pichler, P. Piantanida and G. Matz, “Distributed Information-Theoretic Biclustering”, arXiv:1602.04605.

M.S. Rahman and A. B. Wagner, “On the Optimality of Binning for Distributed Hypothesis Testing”, IEEE Trans. Inform. Theory, vol. 58, no. 10, pp. 6282–6303, October 2012.

N. Slonim and N. Tishby, “Document clustering using word clusters via the information bottleneck method”, in Proc. of 23rd Ann. Int'l ACM SIGIR Conf. on Res. and Dev. in Info. Retrieval, 2000, pp. 208–215.

Y. Steinberg, “Coding and common reconstruction”, IEEE Trans. on Info. Theory, vol. 55, no. 11, pp. 4995–5010, Nov. 2009.

D. Strouse and D. J. Schwab, “The deterministic information bottleneck”, arXiv:1604.00268.

I. Sutskover, S. Shamai and J. Ziv, “Extremes of Information Combining”, IEEE Trans. Inform. Theory, vol. 51, pp. 1313–1325, April 2005.

N. Tishby, F. C. Pereira, and W. Bialek, “The information bottleneck method,” in Proc. 37th Annual Allerton Conf. on Comm., Control, and Computing, 1999, pp. 368–377.

N. Tishby and N. Zaslavsky, “Deep learning and the information bottleneck principle,” Information Theory Workshop (ITW), 2015, 26 April–1 May 2015, Jerusalem, Israel.

K. Viswanatha, E. Akyol and K. Rose, “The Lossy common information of Correlated Sources”, IEEE Trans. Inform. Theory, vol. 60, no. 6, pp. 3238–3253, June 2014.

A. D. Wyner, “The Common Information of Two Dependent Random Variables,” Inf. Theory, vol. 21, no. 2, pp. 163–179, March 1975.

A. Winkelbauer, S. Farthofer, and G. Matz, “The Rate-Information Trade-off for Gaussian Vector Channels”, 2014 IEEE International Symposium on Information Theory, Honolulu, Hawaii, USA, June 29–July 4, 2014.

References (cont.)

H. Witsenhausen and A. Wyner, “A conditional entropy bound for a pair of discrete random variables”, IEEE Trans. Inform. Theory, vol. 21, pp. 493–501, Sep. 1975.

H. Witsenhausen, “Indirect rate distortion problems”, IEEE Trans. IT, vol. 26, pp. 518–521, Sep. 1980.

A. D. Wyner, “On source coding with side information at the decoder”, IEEE Trans. on Info. Theory, vol. 21, pp. 294–300, 1975.

Q. Yang, P. Piantanida and D. Gunduz, “The Multi-layer Information Bottleneck Problem”, arXiv:1711.05102.

H. Yang and R. D. Wesel, “On the Most Informative Boolean Functions of the Very Noisy Channel”, arXiv:1807.11289.

A. Zaidi, I. E. Aguerri, G. Caire and S. Shamai (Shitz), “Uplink Oblivious Cloud Radio Access Networks: An Information Theoretic Overview”, Inform. Theory & Applic. (ITA2018), Feb. 11–16, 2018, San-Diego, USA.

A. Zaidi, I. E. Aguerri, G. Caire and S. Shamai (Shitz), “On the Capacity of Cloud Radio Access Networks with Oblivious Relaying”, arXiv:1701.07237.

Shlomo Shamai

Viterbi Faculty of Electrical Engineering, Technion—Israel Institute of Technology

"Perspectives on Information Bottleneck Problems"

Abstract:

This talk focuses on variants of the bottleneck problem taking an information theoretic perspective. The intimate connections of this setting to: Remote Source-Coding; Information Combining; Common Reconstruction; The Wyner-Ahlswede-Korner Problem; The Efficiency of Investment Information; CEO Source Coding under Log-Loss and others will be highlighted. We discuss the distributed information bottleneck problem with emphasis on the Gaussian model and highlight the basic connections to the uplink Cloud Radio Access Networks (CRAN) with oblivious processing. For this model, the optimal tradeoffs between rates (i.e. complexity) and information (i.e. accuracy) in the discrete and vector Gaussian frameworks is determined. In the concluding outlook, some interesting problems are mentioned such as the characterization of the optimal inputs ('features') distributions under power limitations maximizing the 'accuracy' for the Gaussian information bottleneck, under 'complexity' constraints.

Based on joint works with: A. Zaidi (Universite Paris-Est, Paris).

The research is supported by the European Union's Horizon 2020 Research and Innovation Programme: No. 694630.

Thank you!