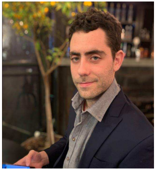
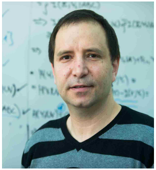
Capacity of Continuous Channels with Memory via Directed Information Neural Estimator

Ziv Aharoni[1], Dor Tsur[1], Ziv Goldfeld[2], Haim H. Permuter[1]

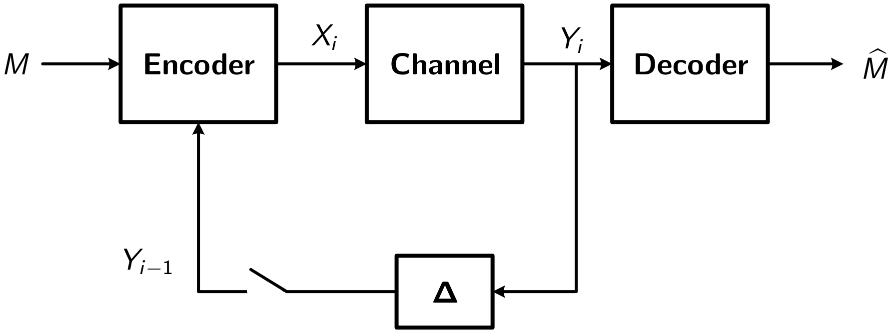


[1]Ben-Gurion University of the Negev

[2]Cornell University

International Symposium on Information Theory
June 21st, 2020

Communication Channel



- Continuous alphabet
- Time invariant channel with memory
- channel is unknown

Capacity

- Feedback **is not** present:

$$C_{\mathrm{FF}} = \lim_{n\to\infty} \sup_{P_{X^n}} \frac{1}{n} I(X^n; Y^n)$$

- Feedback **is** present:

$$C_{\mathrm{FB}} = \lim_{n\to\infty} \sup_{P_{X^n \| Y^{n-1}}} \frac{1}{n} I(X^n \to Y^n)$$

where $I(X^n \to Y^n)$ is the **directed information** (DI)

Capacity

- Feedback **is not** present:

$$C_{\mathrm{FF}} = \lim_{n\to\infty} \sup_{P_{X^n}} \frac{1}{n}\mathbf{I}(\mathbf{X^n} \to \mathbf{Y^n})$$

- Feedback **is** present:

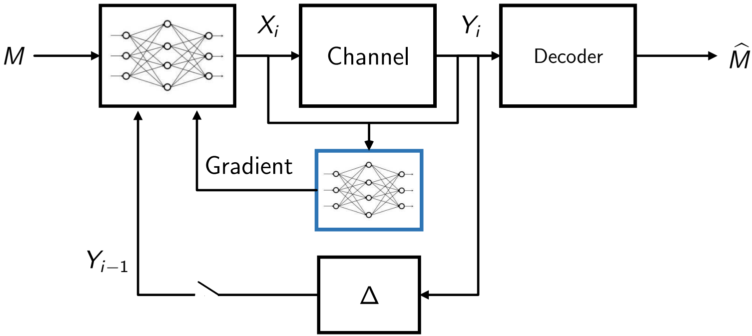
$$C_{\mathrm{FB}} = \lim_{n\to\infty} \sup_{P_{X^n \| Y^{n-1}}} \frac{1}{n}\mathbf{I}(\mathbf{X^n} \to \mathbf{Y^n})$$

where $I(X^n \to Y^n)$ is the **directed information** (DI)

- DI is a unifying measure for feed-forward (FF) and feedback (FB) capacity

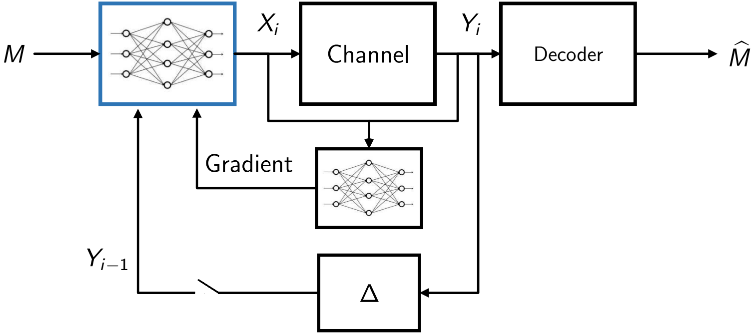
Talk Outline

- Directed Information Neural Estimator (DINE)



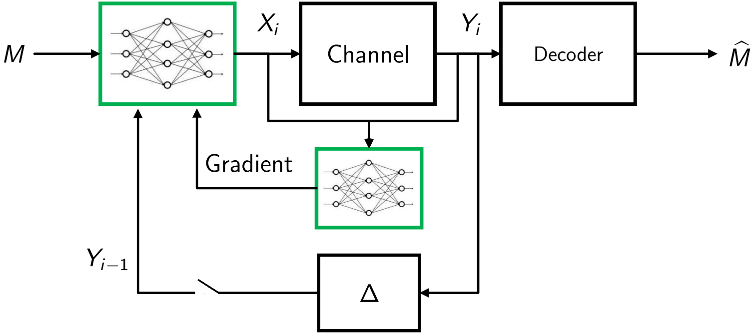
Talk Outline

- Directed Information Neural Estimator (DINE)
- Neural Distribution Transformer (NDT)



Talk Outline

- Directed Information Neural Estimator (DINE)
- Neural Distribution Transformer (NDT)
- Capacity estimation



Preliminaries - Donsker-Varadhan

Theorem (Donsker-Varadhan Representation)

The KL-divergence between the probability measures P and Q, can be represented by

$$D_{\mathrm{KL}}(P \| Q) = \sup_{\mathrm{T}:\Omega \longrightarrow \mathbb{R}} \mathbb{E}_P[\mathrm{T}] - \log \mathbb{E}_Q\left[e^{\mathrm{T}}\right]$$

where, T is measurable and expectations are finite.

- For mutual information:

$$I(X;Y) = \sup_{\mathrm{T}:\Omega \longrightarrow \mathbb{R}} \mathbb{E}_{P_{XY}}[\mathrm{T}] - \log \mathbb{E}_{P_X P_Y}\left[e^{\mathrm{T}}\right]$$

MINE (Y. Bengio Keynoe ISIT '19)

Mutual Information Neural Estimator:

Given $\{x_i, y_i\}_{i=1}^n$

- Approximation

$$\hat{I}(X;Y) = \sup_{\theta \in \Theta} \mathbb{E}_{P_{XY}}[\mathrm{T}_\theta] - \log \mathbb{E}_{P_X P_Y}\left[e^{\mathrm{T}_\theta}\right]$$

- Estimation

$$\hat{I}_n(X,Y) = \sup_{\theta \in \Theta} \frac{1}{n} \sum_{i=1}^n T_\theta(x_i, y_i) - \log \frac{1}{n} \sum_{i=1}^n e^{T_\theta(x_i, \widetilde{y}_i)}$$

Estimator Derivation

- DI as entropies difference

$$I(X^n \to Y^n) = h(Y^n) - h(Y^n \| X^n)$$

where $h(Y^n \| X^n) = \sum_{i=1}^{n} h\left(Y_i | X^i, Y^{i-1}\right)$

- Using an reference measure:

$$I(X^n \to Y^n) = I(X^{n-1} \to Y^{n-1}) + \underbrace{D_{\mathrm{KL}}(P_{Y^n\|X^n} \| P_{Y^{n-1}\|X^{n-1}} \otimes P_{\widetilde{Y}} | P_{X^n})}_{D^{(n)}_{Y\|X}} - \underbrace{D_{\mathrm{KL}}(P_{Y^n} \| P_{Y^{n-1}} \otimes P_{\widetilde{Y}})}_{D^{(n)}_{Y}}$$

$P_{\widetilde{Y}}$ is some uniform i.i.d reference measure of the dataset.

Estimator Derivation

- DI Rate as a difference of KL-divergences:

$$I(X^n \to Y^n) = I(X^{n-1} \to Y^{n-1}) \quad + \quad \underbrace{D^{(n)}_{Y\|X} - D^{(n)}_{Y}}_{\text{increment in info. in step n}}$$

Estimator Derivation

- DI Rate as a difference of KL-divergences:

$$D^{(n)}_{Y\|X} - D^{(n)}_{Y} \xrightarrow{n\to\infty} I(\mathcal{X} \to \mathcal{Y})$$

The limit exists for ergodic and stationary processes

Estimator Derivation

- DI Rate as a difference of KL-divergences:

$$D^{(n)}_{Y\|X} - D^{(n)}_{Y} \xrightarrow{n\to\infty} I(\mathcal{X} \to \mathcal{Y})$$

- **The goal:**

  Estimate $D^{(n)}_{Y\|X}, D^{(n)}_{Y}$

Directed Information Neural Estimator

- Apply DV formula on $D^{(n)}_{Y\|X}, D^{(n)}_{Y}$:

$$\widehat{D}^{(n)}_{Y} = \sup_{T:\Omega\to\mathbb{R}} \mathbb{E}_{P_{Y^n}}\left[\mathrm{T}(Y^n)\right] - \mathbb{E}_{P_{Y^{n-1}}\otimes P_{\tilde{Y}}}\left[\exp\left\{\mathrm{T}(Y^{n-1},\tilde{Y})\right\}\right]$$

where the optimal solution is $\mathrm{T}^* = \log \frac{P_{Y_n|Y^{n-1}}}{P_{\tilde{Y}}}$

Directed Information Neural Estimator

- Approximate T with a recurrent neural network (RNN)

$$\widehat{D}_Y^{(n)} = \sup_{\theta_{\mathbf{Y}}} \mathbb{E}_{P_{Y^n}}\left[\mathrm{T}_{\theta_{\mathbf{Y}}}(Y^n)\right] - \mathbb{E}_{P_{Y^{n-1}} \otimes P_{\tilde{Y}}}\left[\exp\left\{\mathrm{T}_{\theta_{\mathbf{Y}}}(Y^{n-1}, \tilde{Y})\right\}\right]$$

Directed Information Neural Estimator

- Estimate expectations with empirical means

$$\widehat{D}_Y^{(n)} = \sup_{\theta_Y} \frac{1}{n} \sum_{i=1}^{n} \mathsf{T}_{\theta_Y}(y_i|y^{i-1}) - \log \left(\frac{1}{n} \sum_{i=1}^{n} e^{\mathsf{T}_{\theta_Y}(\widetilde{y}_i|y^{i-1})} \right)$$

Directed Information Neural Estimator

- Estimate expectations with empirical means

$$\widehat{D}_Y^{(n)} = \sup_{\theta_Y} \frac{1}{n}\sum_{i=1}^{n} \mathrm{T}_{\theta_Y}(y_i|y^{i-1}) - \log\left(\frac{1}{n}\sum_{i=1}^{n} e^{\mathrm{T}_{\theta_Y}(\widetilde{y}_i|y^{i-1})}\right)$$

Finally, $\quad \widehat{I}^{(n)}(\mathcal{X} \to \mathcal{Y}) = \widehat{D}_{XY}^{(n)} - \widehat{D}_Y^{(n)}$

Consistency

Theorem (DINE consistency)

Let $\{X_i, Y_i\}_{i=1}^{\infty} \sim \mathbb{P}$ *be jointly stationary ergodic stochastic processes. Then, there exist RNNs* $\mathsf{F}_1 \in \mathsf{RNN}_{d_y,1}, \mathsf{F}_2 \in \mathsf{RNN}_{d_{xy},1}$*, such that DINE* $\widehat{I}_n(\mathsf{F}_1, \mathsf{F}_2)$ *is a strongly consistent estimator of* $I(\mathcal{X} \to \mathcal{Y})$*, i.e.,*

$$\lim_{n\to\infty} \widehat{I}_n(\mathsf{F}_1, \mathsf{F}_2) \stackrel{a.s}{=} I(\mathcal{X} \to \mathcal{Y})$$

Consistency

Theorem (DINE consistency)

Let $\{X_i, Y_i\}_{i=1}^{\infty} \sim \mathbb{P}$ *be jointly stationary ergodic stochastic processes. Then, there exist RNNs* $\mathrm{F}_1 \in \mathrm{RNN}_{d_y,1}, \mathrm{F}_2 \in \mathrm{RNN}_{d_{xy},1}$*, such that DINE* $\widehat{I}_n(\mathrm{F}_1, \mathrm{F}_2)$ *is a strongly consistent estimator of* $I(\mathcal{X} \to \mathcal{Y})$*, i.e.,*

$$\lim_{n\to\infty} \widehat{I}_n(\mathrm{F}_1, \mathrm{F}_2) \overset{a.s}{=} I(\mathcal{X} \to \mathcal{Y})$$

Sketch of proof:

- Represent the solution T^* by a dynamic system.
- Universal approximation of dynamical system with RNNs.
- Estimation of expectations with empirical means.

Implementation

$$\widehat{D}_Y^{(n)} = \sup_{\theta_Y} \frac{1}{n}\sum_{i=1}^{n} \mathsf{T}_{\theta_Y}(y_i|y^{i-1}) - \log\left(\frac{1}{n}\sum_{i=1}^{n} e^{\mathsf{T}_{\theta_Y}(\widetilde{y}_i|y^{i-1})}\right)$$

Implementation

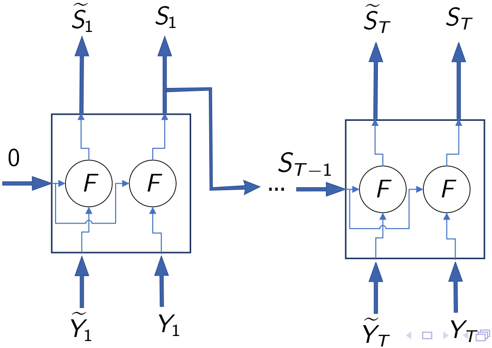
$$\widehat{D}_Y^{(n)} = \sup_{\theta_Y} \frac{1}{n}\sum_{i=1}^{n} \mathsf{T}_{\theta_Y}(y_i|y^{i-1}) - \log\left(\frac{1}{n}\sum_{i=1}^{n} e^{\mathsf{T}_{\theta_Y}(\widetilde{y}_i|y^{i-1})}\right)$$

Adjust RNN to process both inputs and carry the state generated by true samples

Implementation

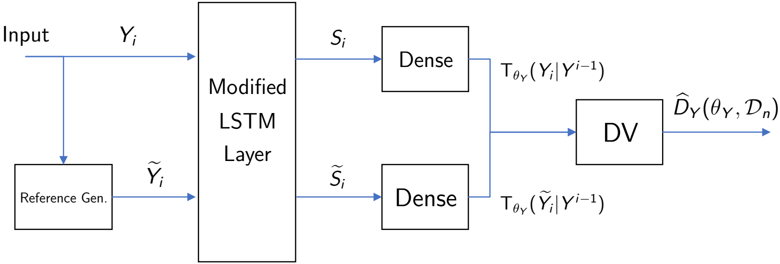
$$\widehat{D}_Y^{(n)} = \sup_{\theta_Y} \frac{1}{n}\sum_{i=1}^{n} \mathrm{T}_{\theta_Y}(y_i|y^{i-1}) - \log\left(\frac{1}{n}\sum_{i=1}^{n} e^{\mathrm{T}_{\theta_Y}(\widetilde{y}_i|y^{i-1})}\right)$$

Adjust RNN to process both inputs and carry the state generated by true samples

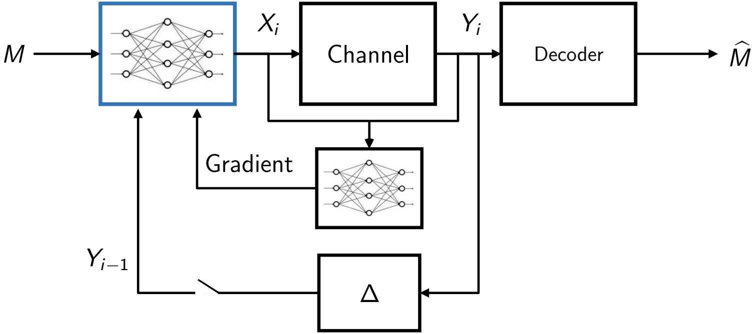


Implementation

- Complete system layout for the calculation of $\widehat{\mathcal{D}}_Y^{(n)}$



Neural Distribution Transformer (NDT)



NDT

- Model M as i.i.d Gaussian noise $\{N_i\}_{i\in\mathbb{Z}}$.
- The NDT a mapping

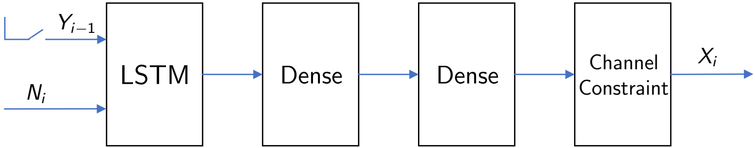
$$\begin{array}{ll} \textbf{w/o feedback:} & \mathrm{NDT} : N^i \longmapsto X_i \\ \textbf{w/ feedback:} & \mathrm{NDT} : N^i, Y^{i-1} \longmapsto X_i \end{array}$$

NDT

- Model M as i.i.d Gaussian noise $\{N_i\}_{i\in\mathbb{Z}}$.
- The NDT a mapping

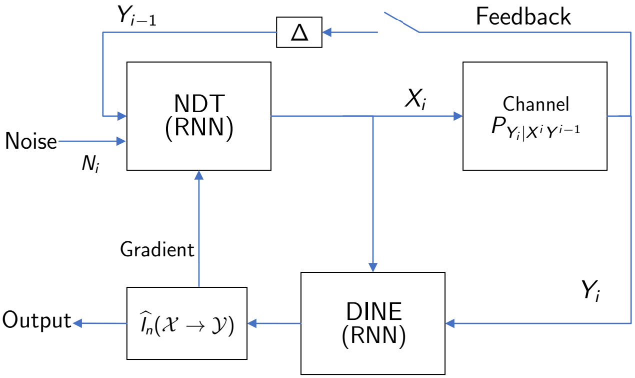
$$\begin{array}{ll} \textbf{w/o feedback:} & \mathrm{NDT}: N^i \longmapsto X_i \\ \textbf{w/ feedback:} & \mathrm{NDT}: N^i, Y^{i-1} \longmapsto X_i \end{array}$$

- NDT is modeled by an RNN



Capacity Estimation

- Iterating between DINE and NDT.



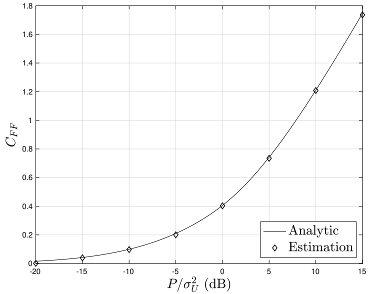
Results

- Channel - MA(1) additive Gaussian noise (AGN):

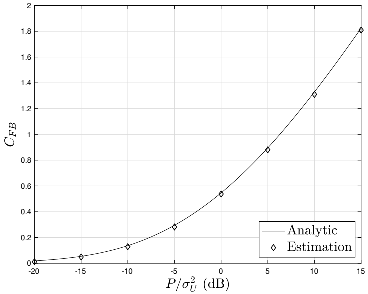
$$Z_i = \alpha U_{i-1} + U_i$$
$$Y_i = X_i + Z_i$$

where, $U_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$, X_i is the channel input sequence bound to the power constraint $\mathbb{E}\left[X_i^2\right] \leq P$, and Y_i is the channel output.

Estimation performance



(a) Feed-forward Capacity



(b) Feedback Capacity

Conclusion and Future Work

Conclusions:

- Estimation method for both FF and FB capacity.
- **Pros:** mild assumptions on the channel
- **Cons:** lack of provable bounds

Conclusion and Future Work

Conclusions:

- Estimation method for both FF and FB capacity.
- **Pros:** mild assumptions on the channel
- **Cons:** lack of provable bounds

Future Work:

- Generalize for more complex scenarios (e.g multi-user)
- Obtain provable bounds on fundamental limits

Conclusion and Future Work

Conclusions:

- Estimation method for both FF and FB capacity.
- **Pros:** mild assumptions on the channel
- **Cons:** lack of provable bounds

Future Work:

- Generalize for more complex scenarios (e.g multi-user)
- Obtain provable bounds on fundamental limits

Thank You!