Source Coding With Delayed Side Information

Osvaldo Simeone\textsuperscript{(1)} and Haim Permuter\textsuperscript{(2)}

\textsuperscript{(1)}New Jersey Institute of Technology (NJIT), USA,
\textsuperscript{(2)}Ben-Gurion University of the Negev, Israel

ISIT 2012
"Non-causal" reconstruction [Gray, 72]
"Causal" reconstruction with delay $d$

Computational complexity, delay and real-time constraints

Short treatment in [Venkataramanan and Pradhan, 05]
"Causal" reconstruction with delay $d$

With $X_i = Y_i$: source coding with feedforward [Weissman and Merhav, 03]
Previous Work

Memoryless sequences $X^n$ and $Y^n$
- $d = 0$ (causal state information) ↔ non-causal side information ($\hat{X}_i(M, Y^n)$) [Gray, 72]

$$R(D) = \min_{p(\hat{x}|x,y): \mathbb{E}[d(X,\hat{X})] \leq D} I(X; \hat{X}|Y)$$
Memoryless sequences $X^n$ and $Y^n$
- $d > 0 \iff$ no side information ($\hat{X}_i(M)$), but can simplify encoding
  [Weissman and Merhav, 03] [Martinian and Wornell, 04] [Pradhan, 04]

$$R(D) = \min_{p(\hat{x}|x): \mathbb{E}[d(X,\hat{X})] \leq D} I(X;\hat{X})$$
Previous Work

Sequences $X^n$ and $Y^n$ with memory with $X_i = Y_i$ (feedforward)
- Feedforward generally useful [Weissman and Merhav, 03]
- Not useful if rate-distortion function equals the Shannon lower bound [Weissman and Merhav, 03]
Previous Work

Sequences $X^n$ and $Y^n$ with memory with $X_i = Y_i$ (feedforward)
- Stationary and ergodic sources [Venkataramanan and Pradhan, 05]
  [Naiss and Permuter, 11]

$$R(D) = \lim_{n \to \infty} \min_{p(\hat{X}^n | X^n) : \frac{1}{n} \sum_{i=1}^{n} I(X^n_{i-d+1}; \hat{X}_i|\hat{X}^{i-1}, X^{i-d})} \frac{1}{n} I(\hat{X}^{n-d+1} \rightarrow X^n)$$

- Algorithm for numerical evaluation [Naiss and Permuter, 11]
Sequences $X^n$ and $Y^n$ with memory and $X_i \neq Y_i$:
- Stationary and ergodic sources [Venkataramanan and Pradhan, 05]

$$R(D) = \inf \{p(\hat{x}^n|x^n,y^n)\}: \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} I(X^n, Y_{i-d+1}; \hat{X}_i|\hat{X}^{i-1}, Y^{i-d})$$

$$\lim \sup_{n \to \infty} \frac{\sum_{i=1}^{n} I(X^n, Y_{i-d+1}; \hat{X}_i|\hat{X}^{i-1}, Y^{i-d})}{n \leq D}$$
Contributions

- Hidden Markov Model
- Single-letter characterization for minimum rate required for lossless reconstruction (any $d \geq 0$)

- Rate-distortion function for $d = 0, 1$
- Binary example
$Y^n$ is a stationary ergodic Markov chain

$X_i \sim q(x|y_i)$ if $Y_i = y_i$ (hidden Markov model)

$$p(x^n, y^n) = \pi(y_1) \prod_{i=2}^{n} w_1(y_i|y^{i-1}) q(x_i|y_i)$$
$Y^n = (Y_1, \ldots, Y_n) \rightarrow Y^{i-d} = (Y_1, \ldots, Y^{i-d})$

$X^n = (X_1, \ldots, X_n) \rightarrow M \text{ of } nR \text{ bits} \rightarrow \hat{X}_i(M, Y^{i-d})$

- $Y^n$ physical source, measured at the encoder
- $X^n$ symbol-by-symbol processed version of $Y^n$
- $R$ [bits/source symbol]
An \((d, n, R, D_1, D_2)\) code is defined by: 

(i) An encoder function 

\[ f: (\mathcal{X}^n \times \mathcal{Y}^n) \rightarrow [1, 2^{nR}] \]

(ii) a sequence of decoding functions 

\[ g_i: [1, 2^{nR}] \times \mathcal{Y}^{i-d} \rightarrow \mathcal{Z} \]

Distortion constraint 

\[ \frac{1}{n} \sum_{i=1}^{n} E[d(X_i, Y_i, Z_i)] \leq D \]

Standard definition for rate-distortion function \(R_d(D_1, D_2)\)
Lossless Reconstruction

- Hamming distortion metric $d_1(x, y, z_1) = 1(x \neq z_1)$

**Theorem**

*For any delay $d \geq 0$, the rate-distortion function under Hamming distortion is given at $D_1 = 0$ by*

$$R_d(0) = H(X_{d+1} | X_d^d, Y_1),$$

*where the conditional entropy is calculated with respect to the distribution*

$$p(y_1, x_1) = \pi(y_1)q(x_1 | y_1) \text{ for } d = 0,$$

*and $p(y_1, x_2, \ldots, x_{d+1}) = \pi(y_1)$*

$$\sum_{y_i \in \mathcal{Y}} \prod_{i=2}^{d+1} w_i(y_i | y_{i-1})q(x_i | y_i), \text{ for } d \geq 1.$$
Comments

- $H(X_{d+1} | X_2^d, Y_1)$:

![Diagram showing $Y_1$, $X_2^d$, $X_{d+1}$ with delay $d$]

- Ex: $R_0(0) = H(X_1 | Y_1)$ (as with non-causal side information)
- Ex: $R_1(0) = H(X_2 | Y_1)$
- Ex: $R_\infty(0) = H(\mathcal{X})$ (entropy rate)
• Delayed side information is generally useful

\[ R_d(0) = H(X_{d+1} | X_2^d, Y_1) \leq R_\infty(0) = H(X) \]

• Equality if \( X_i \) is an i.i.d. process or with feedforward (\( X_i = Y_i \))

• Can be obtained from the multi-letter expression in [Venkataramanan and Pradhan 05]
Proof

Partition the interval \([1, n]\) into \(|\mathcal{X}|^{d-1}|\mathcal{Y}|\) subintervals \(\mathcal{I}(\tilde{x}^{d-1}, \tilde{y}) \subseteq [1, n]\), for all \(\tilde{x}^{d-1} \in \mathcal{X}^{d-1}\) and \(\tilde{y} \in \mathcal{Y}\):

\[
\mathcal{I}(\tilde{x}^{d-1}, \tilde{y}) = \{i: i \in [1, n] \text{ and } y_{i-d} = \tilde{y}, \ x_{i-d+1}^{i-1} = \tilde{x}^{d-1}\}.
\]
**Ex: \( d = 2 \)**

\[
\begin{align*}
x^n & = 0,0,1,0,1,0,1,0,1,1 \\
y^n & = 0,1,1,0,1,1,0,0,1,1
\end{align*}
\]

<table>
<thead>
<tr>
<th>( \tilde{x} )</th>
<th>( \tilde{y} )</th>
<th>( \mathcal{I}(\tilde{x}, \tilde{y}) )</th>
<th>( x^\mathcal{I}(\tilde{x}, \tilde{y}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>{1, 2, 3, 9}</td>
<td>[0, 0, 1, 1]</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>{5, 7}</td>
<td>[1, 1]</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>{6, 10}</td>
<td>[0, 1]</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>{4.8}</td>
<td>[0, 0]</td>
</tr>
</tbody>
</table>
By the ergodicity of process $X_i$ and $Y_i$, $x^{I}(\tilde{x}^{d-1}, \tilde{y}_1)$ is of length less than $npY_1X_2,...,X_d(\tilde{y}, \tilde{x}^{d-1}) + \epsilon$.

Each sequence $X^{I}(\tilde{x}^{d-1}, \tilde{y})$ i.i.d. with distribution $p_{X_{d+1}|Y_1X_2,...,X_d}(\cdot | \tilde{y}, \tilde{x}^{d-1})$.

Entropy coding: $H(X_{d+1}|X_2^d = \tilde{x}^{d-1}, Y_1 = \tilde{y}) + \epsilon$ bits/source symbol.
Recover the individual sequences $x^{T(\tilde{x}^{d-1}, \tilde{y})}$ by decoding of the entropy codes

Need to reorder the symbols to obtain $x^n$

Can be done since at time $i$, the decoder knows $Y_{i-d}$ and the previously decoded $X^{i-1} \rightarrow I(\tilde{x}^{d-1}, \tilde{y})$ to which the current symbol $X_i$ belongs
The proof can be generalized to show that

\[ R_d(0) = \lim_{n \to \infty} \frac{1}{n} H(X^n \| Y^{n-d}) \]

\[ = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} H(X_i | X_i^{i-1}, Y_i^{i-d}) \]

Rate decrease due to delayed side information

\[ \lim_{n \to \infty} \frac{1}{n} H(X^n) - H(X^n \| Y^{n-d}) \]

\[ = \lim_{n \to \infty} \frac{1}{n} I(Y^{n-d} \to X^n) \]

See [Kim et al '11] for other interpretations of directed mutual information
When the Side Information May Be Delayed

\[ X^n \xrightarrow{\text{Enc}} Y^n \]

Enc \hspace{3cm} \text{Dec 1} \hspace{3cm} \text{Dec 2}

\[ \hat{X}_{1i}(M, Y^{i-d}) \]

\[ \hat{X}_{2i}(M, Y^i) \]
Lossy Reconstruction

**Theorem**

For any delay \( d \geq 0 \) and distortion pair \((D_1, D_2)\), the following rate is achievable

\[
R_d^{(a)}(D_1, D_2) = \min I(XY; \hat{X}_1 | Y_d) + I(X; \hat{X}_2 | YY_d \hat{X}_1)
\]

with mutual informations evaluated with respect to the joint distribution

\[
p(x, y, y_d, z_1, z_2) = \pi(y_d) w_d(y | y_d) q(x | y) p(\hat{x}_1, \hat{x}_2 | x, y, y_d),
\]

and where the minimum is taken over distributions \( p(\hat{x}_1, \hat{x}_2 | x, y, y_d) \) such that

\[
E[d_j(X, Y, \hat{X}_j)] \leq D_j, \text{ for } j = 1, 2.
\]

Moreover, the above is the rate-distortion function, i.e.,

\[
R_d^{(a)}(D_1, D_2) = R_d(D_1, D_2), \text{ for } d = 0 \text{ and } d = 1.
\]
Achievability using similar demux operations and conventional rate-distortion codes:
- $I(XY; \hat{X}_1|Y_d)$ bits/source sample to inform decoder 1 (and decoder 2) about $Z_1^n$
- $I(X; \hat{X}_2|Y Y_d \hat{X}_1)$ bits/source sample to inform decoder 2 about $Z_2^n$
- $Y_i$ binary Markov chain with symmetric transition probabilities
  \[ w_1(1|0) = w_1(0|1) \triangleq \varepsilon \]
- $X_i = Y_i \oplus N_i$, $N_i \sim Ber(q)$, $q \leq 0.5$
- Hamming distortion $d(x, \hat{x}) = x \oplus \hat{x}$
Example: Lossless Reconstruction

- $q = 0.1$

Graph showing $R_d(0)$ [bits/source symbol] vs. $d$ for different values of $\epsilon$: $\epsilon = 0.1, 0.2, 0.3, 0.4, 0.5$.
Example: Lossless Reconstruction

- $\varepsilon = 0.1$

\[ R_d(0) \text{ [bits/source symbol]} = \begin{cases} \vdots & \text{for } d = 0 \\ \vdots & \text{for } d \to \infty \end{cases} \]

- $X = Y$ (Markov, feedfoward)
- $X$ indep of $Y$ (i.i.d.)

\[ \text{no side information } H(X) \quad (d \to \infty) \]
Example: Lossless Reconstruction

- $\varepsilon = 0.1$

Graph showing $R_d(0)$ [bits/source symbol] vs $q$ with curves for different values of $d$. The graph includes labels for $H(X)$ and $H(X_1)$, and notes that $X = Y$ (Markov, feedfoward) and $X$ indep of $Y$ (i.i.d.).
Example: Lossy Reconstruction

- Rate-distortion function for $d = 0, 1$

$$R_d^{(a)}(D_1) = H_b(\varepsilon^{(d)} \ast q) - H_b(D_1)$$

for $0 \leq D_1 \leq \min\{\varepsilon^{(d)} \ast q, 1 - \varepsilon^{(d)} \ast q\}$ and $R_d^{(a)}(D_1) = 0$ otherwise, where $p \ast q \triangleq p(1 - q) + (1 - p)q$

- For feedforward and $d = 1$, it recovers [Weissman and Merhav, 03] (rate-distortion of the innovation): for $0 \leq D_1 \leq \min\{\varepsilon, 1 - \varepsilon\}$

$$R_1(D_1) = H_b(\varepsilon) - H_b(D_1) \leq R_{\text{without feedforward}}(D_1)$$

[Gray 70]

- if $D_1$ larger than a critical value, feedforward, unlike in the lossless case, can be useful
Concluding Remarks

- Compressing information sources in the presence of delayed side information
- Applications: sensor networks and prediction/denoising
- General multiletter characterization [Venkataramanan and Pradhan, 05]
- Focusing on a specific class of (hidden Markov model) sources, single-letter characterizations
- Simple achievable schemes based on standard “off-the-shelf” compression techniques
- Extension to unknown delay