HD Computing in Communication

Ping-Chen Huang
Nov. 17, 2021
Outline

• Overview of communication problems
• Principles related to HD computing in communication
  • Law of large numbers & random codes
  • Orthogonality via random projection
• Applications of HD computing in communication
• AI/Machine learning in communication
Main Ideas from Shannon’s paper

- Treat the information source as stochastic
- Separate the transmitter/receiver design into two sub-problems – “source-channel separation”

Why the madness of having two competing encoders?
Recall Module 7 (Fritz’s lecture)

Mapping data to vector spaces

Source coding (remove redundancy in data)

<table>
<thead>
<tr>
<th>Data lie in subspace (SS)</th>
<th>Learning method</th>
<th>Coordinates in SS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear low-D SS</td>
<td>PCA</td>
<td>Axes of covariance matrix</td>
</tr>
<tr>
<td>Nonlinear low-D SS Clusters</td>
<td>Manifold learning</td>
<td>location on manifold</td>
</tr>
<tr>
<td>Union of lin. low-D SS</td>
<td>Sparse coding</td>
<td>Axes of Indep. Comp.</td>
</tr>
<tr>
<td>Union of nonlin. Low-D SS</td>
<td>Manifold learning</td>
<td>Manifold number + loc.</td>
</tr>
</tbody>
</table>

Vector encoding of the new coordinates

Feature local: a neuron’s activity encodes a coordinate: PCA, ICA,…

Distributed: values of a coordinate are represented by many neurons: VSA
Main ideas from Shannon’s paper (cont’d)

• Nice answers to both questions:
  i. Min. bits to represent a source = H (source entropy) bits/symbol
  ii. Max. rate of reliable communication over a channel = C (capacity) bits/use

• Source can be reliably communicated over a channel iff $H < C$
Surprise from this theory

• We can communicate with arbitrarily small error rates without operating at zero-rate

Repetition code

Capacity achieving code

Eg. for binary symmetric channel with bit flip probability $p$
Open questions from Shannon’s paper

• How to design codes?
  ⇒ Source codes: Huffman, Lemple-Ziv (77, 78)
  ⇒ Channel codes: Turbo, LDPC, Polar, etc.

• Network Communication?
  ⇒ Distributed source compression: Slepian-Wolf Theory
  ⇒ Distributed communication
    • Multiple-Access Channel (MAC)
    • Broadcast Channel (BC)
    • General network communication
Elements in digital communication systems

Texts, images, audio, etc.

BPSK, 16-QAM, FSK, OFDM, etc.

BEC, BSC
AWGN
Band-limited Fading

Information source and input transducer
Source encoder
Channel encoder
Digital modulator

Channel

Output transducer
Source decoder
Channel decoder
Digital demodulator
Entropy and mutual information

• The entropy of a discrete random variable

\[ H(X) = - \sum_{x \in X} p(x) \log p(x) \]

• The joint entropy of a pair of discrete random variables \((X,Y)\) with joint distribution \(p(x,y)\)

\[ H(X,Y) = - \sum_{x \in X} \sum_{y \in Y} p(x,y) \log p(x,y) \]

• Mutual information is the relative entropy between \(p(x,y)\) and \(p(x)p(y)\)

\[
I(X;Y) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}
\]

\[
= H(X) + H(Y) - H(X,Y)
\]

\[
= H(Y) - H(Y|X)
\]
Basic Channel Models

Binary Erasure Channel (BEC)

$$C = 1 - p \quad \text{[bits/use]}$$

Binary Symmetric Channel (BSC)

$$C = 1 - H(p) \quad \text{[bits/use]}$$

Additive White Gaussian Noise Channel (AWGN)

$$C' = \frac{1}{2} \log(1 + \frac{p}{\sigma^2}) \quad \text{[bits/use]}$$
Asymptotic Equipartition Property (AEP)

• The analog of the law of large numbers in information theory:

\[
\frac{1}{n} \log p(X_1, X_2, ..., X_n) \to H(X) \quad \text{in probability.}
\]

• AEP holds when the source has memory (Markov source)
Typical Sequence

- Example: $X=\{a, b\}$, $p(a) = p$, $p(b) = 1-p$, where $p \neq 0.5$

Look at blocks of length $n$, we expect roughly $np$ $a$’s and $n(1-p)$ $b$’s in the sequences will occur with highest probability $\Rightarrow$ typical sequence $X$

$$p(x) = p^{np} (1 - p)^{n(1-p)} = 2^{-nH(X)}$$

Min # bits to represent the source:

$$\frac{nH(X)}{n} = H(X) \text{ bits/symbol}$$
Channel Coding Setup

- $W$ is uniform in $\{1, 2, \cdots, 2^{nR}\}$
  - $n$: block length
  - $R$: rate [bits/usage]
- Up to which $R$ can we still reliably communicate over the channel?

**Channel Coding Theorem:**

For any $R < C$, there exists a sequence of $(2^{nR}, n)$ codes, such that

$$Pr(W \neq \hat{W}) \to 0 \text{ as } n \to \infty$$
Proof of the Channel Coding Theory using random codebooks and the joint typical decoder

- Coder: i.i.d. randomly chosen according to $p(x)$
- Decoder: Choose unique $w$ such that $(X^n(w), Y^n) \in A^{(n)}_\epsilon$

$$Pr(A^{(n)}_\epsilon) \to 1 \text{ as } n \to \infty$$
Proof of the Channel Coding Theory (cont’d)

• Assume $W=1$ is transmitted, there are two error events:

  • $E_0 := (X^n(1), Y^n) \notin A^{(n)}_\epsilon$

  • $E_i := (X^n(i), Y^n) \in A^{(n)}_\epsilon$ for $i \geq 2$

\[
Pr(\mathcal{E}|W = 1) = Pr(E_0 \cup E_1 \cup \ldots \cup E_{2^n R}|W = 1) \\
\leq Pr(E_0|W = 1) + \sum_{i=2}^{2^n R} Pr(E_i|W = 1)
\]

\[
\frac{2^{n H(X,Y)}}{2^{n (H(X) + H(Y))}} = 2^{-n I(X;Y)}
\]
Proof of the Channel Coding Theory (cont’d)

- \( Pr(\mathcal{E}|W = 1) \leq (2^{nR} - 1) \cdot 2^{-nI(X;Y)} \)

\( \Rightarrow \) If \( R < \max_{p(x)} I(X;Y) := C \), then \( Pr(\mathcal{E}) \to 0 \) as \( n \to \infty \).

- \( P(\mathcal{E}) = E_C(Pr(\mathcal{E}|\mathcal{C})) \)

\( \Rightarrow \forall n, \exists \mathcal{C}_n \) such that \( Pr(\mathcal{E}|\mathcal{C}_n) \to 0 \)

The proof is non-constructive – it shows a good code exists, but does not give the explicit encoding scheme.
Random Codes in Reality

• **Linear block codes** with random generator matrix
  
  - \( x = uG \)
  
  - \( G_{ij} \in \{0, 1\} \) i.i.d. with 0 and 1 equally likely

• For BEC and BSC channels, there exists a \( G_n \) that approaches capacity.

• Although non-structured \( G \) is easy to encode, it’s hard to decode:
  
  - \( \min \|uG - y\| \)
  
  - Decoding complexity \( \sim O(n^3) \) (NP-hard)
Random Codes in Reality

**Fountain codes** are a new class of random linear codes with
- Sparse generator matrix
- Decoding complexity $O(n \log(n/\delta))$ with decoding probability $>1-\delta$

**Approaches capacity for erasure channels with unknown erasure probabilities**
- Sender sends a stream of encoded bits.
- Receivers collect bits until they are reasonably sure that they can recover the content from the received bits, and send STOP feedback to sender.
- **Automatic adaptation**: Receivers with larger loss rate need longer to receive the required information.

**Practical realizations**
- L-T code
- Raptor code
- Online code
Remarks about channel codes

• Important considerations
  • Capacity achieving
  • Efficient decoding
    ⇒ Find codes with good structures of G (e.g. LDPC code, R-S code)
    ⇒ Suboptimal decoding schemes (iterative decoding)

• Current 5G standard
  • LDPC code + iterative message passing (decoding complexity \(O(n^2)-O(n)\))
  • Polar code + successive cancellation (decoding complexity \(O(n \log n)\))

• HD computing offers error correction, but its power lies beyond this
Outline

• Overview of communication problems
• Principles related to HD computing in communication
  • Law of large numbers & random codes
  • Orthogonality via random projection
• Applications of HD computing in communication
• Machine learning in communication
Direct Sequence Spread Spectrum

- Applications
  - Overcome jammers
  - Message privacy
  - Multiple access (CDMA)

Bandwidth expansion factor (processing gain)

\[
\frac{W}{R} = \frac{T_s}{T_c}
\]

\[
\frac{E_b}{J_0} = \frac{2W/R}{J_{av}/P_{av}}
\]
Frequency Hopping Spread Spectrum

https://www.slideshare.net/HILDA519/spread-spectrum-modulation

[Credit: Proakis]
Capacity of the multiple access channel

For Gaussian channels

\[ Y = X_1 + X_2 + Z \text{ where } Z \sim N(0, N) \]

\[
R_1 \leq I(X_1; Y|X_2) \leq \frac{1}{2} \log(1 + \frac{P_1}{N})
\]

\[
R_2 \leq I(X_2; Y|X_1) \leq \frac{1}{2} \log(1 + \frac{P_2}{N})
\]

\[
R_1 + R_2 \leq I(X_1, X_2; Y) \leq \frac{1}{2} \log(1 + \frac{P_1 + P_2}{N})
\]
Outline

• Overview of communication problems
• principles related to HD computing in communication
  • Law of large numbers & random codes
  • Orthogonality via random projection
• Applications of HD computing in communication
• Machine learning in communication
Hyperdimensional Modulation

Hyperdimensional Modulation

SNR and $\text{Eb}/N_0$

- **Continuous-time AWGN channel**
  - Signal power $= P$
  - Noise PSD $= N_0$
  - Bandwidth $= W$
  - $\text{SNR} = P/N_0W$

- **Discrete-time AWGN channel**
  - Energy per symbol $E_s = P/W$
  - Noise energy = variance of noise $= N_0$
  - SNR in discrete time $= E_s/N_0 = P/N_0W = \text{SNR in continuous time}$

- $\text{Eb}/N_0 = \text{SNR}/R = (E_s/N_0)/(# \text{ of transmissions}/# \text{ of info bits})$
Application to sensing problems – Example 1

• Assign a binary vector to each temp range: eg. \( v_0: [0-10] \), \( v_1: [10-20] \), \( v_2: [20-30] \), …

• Receiver receives the sum of the vectors sent by all sensors \( s = s_1 + s_2 + … \).

• Binary Query: any inventory in a temp range \( i \)?

\[ s \sim v_i \]

• Proportion query: How much inventory is in temp range \( i \)?

\[ z(a_1 v_1 + a_2 v_2 + \cdots + a_M v_M) = s \]

Linear least square regression to find \( a_i \)

Process the single sum vector \( s \) without decoding individual sensor’s reading

Application to sensing problems – Example 2

\[ u(x_1), v(x_1) \]
\[ u(x_2), v(x_2) \]
\[ u(x_3), v(x_3) \]

\[ u(x_1) x^{x_1} \odot u + v(x_1) x^{x_1} \odot v \]
\[ u(x_2) x^{x_2} \odot u + v(x_2) x^{x_2} \odot v \]
\[ u(x_3) x^{x_3} \odot u + v(x_3) x^{x_3} \odot v \]

\[ s = \sum_{x_i} u(x_i) x^{x_i} \odot u + v(x_i) x^{x_i} \odot v \]

- \( x_i \) is the position id in integers
- \( u(x_i) \) and \( v(x_i) \) are different measurements (eg. sound or gas sensor readings)
- \( \mathbf{x}, \mathbf{u}, \mathbf{v} \) are complex HD vectors for position, u sensor, v sensor
Application to sensing problems – Example 2

Patterns of 16 Sensors’ Readings

2nd Pattern (original)

[0 0 0 1 0 1 0 1 0 0 1 0 0 1]

2nd Pattern (R shift +1)

[1 0 0 0 1 0 1 0 1 0 0 1 0 0]

2nd Pattern (R shift +2)

[0 1 0 0 0 1 0 1 0 1 0 0 1 0]
Outline

• Overview of communication problems
• Principles related to HD computing in communication
  • Law of large numbers & random codes
  • Orthogonality via random projection
• Applications of HD computing in communication
• AI/Machine learning in communication
Communication system as an autoencoder

Loss function to minimize: block level error rate

Communication system as an autoencoder


![Figure 4: Constellations produced by autoencoders using parameters (n, k): (a) (2, 2) (b) (2, 4), (c) (2, 4) with average power constraint, (d) (7, 4) 2-dimensional t-SNE embedding of received symbols.](image-url)
Learning to decode

Learning to decode

- Structured codes are easier to learn than random codes

Learning to decode

- Provide a subset of the entire set during training (p%)
- Able to generalize to codewords that it has never seen during training for structure codes, but not for random codes

Learning to construct codes

Conclusions

• The power of HD computing lies in encoding structures, learning, and classification.

• New applications drive communication from traditional point-to-point transmission paradigm into a variety of new problems

⇔ Opportunities for HD computing

Acknowledgements: Connor Bybee, Denis Kleyko, Chris Kymn
References