Prediction-Aided Sequential Communication of Individual Sequences with Distortion Guarantees

Giuseppe Durisi

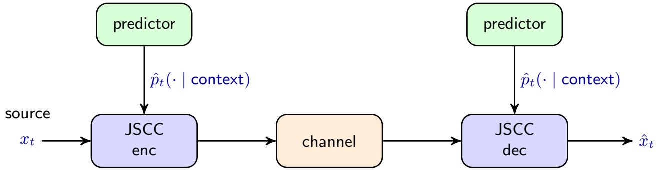
Chalmers, Sweden

Mar., 2026

Joint work with M. Zecchin, U. K. Ganesan, P. Popovski, O. Simeone

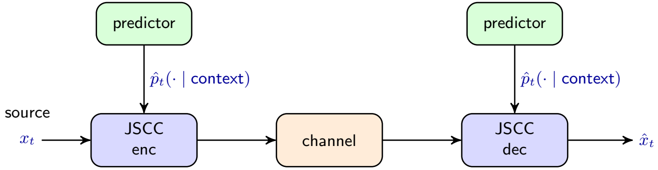
CHALMERS

Prediction-Aided Communication



- Motivations:
  - Connected intelligence vision: wireless devices and network nodes equipped with enough computational power for ML algorithms
  - Success of LLMs as universal source compressors
  - Data-driven approaches to joint-source channel coding (JSCC)
- Applications: efficient transmission of control information, real-time monitoring, collaborative robotics
- Requirements: sequential/zero-delay, universal guarantees over sequences and predictors

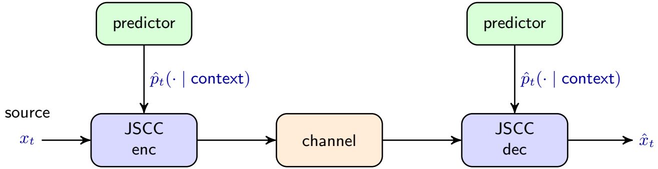
Related Problems



- Fixed-slope universal lossy data compression [Yang, Zhang, Berger '97]

$$\hat{x}^n = \arg\min_{\tilde{x}^n} \mathrm{LZ}(\tilde{x}^n) + s d(x^n, \tilde{x}^n)$$

Related Problems

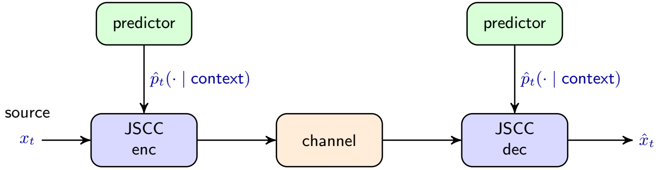


- Fixed-slope universal lossy data compression [Yang, Zhang, Berger '97]

$$\hat{x}^n = \arg\min_{\tilde{x}^n} \mathrm{LZ}(\tilde{x}^n) + s d(x^n, \tilde{x}^n)$$

- Universal zero-delay JSCC [Matloub & Weissman '06]: regret analysis

Related Problems

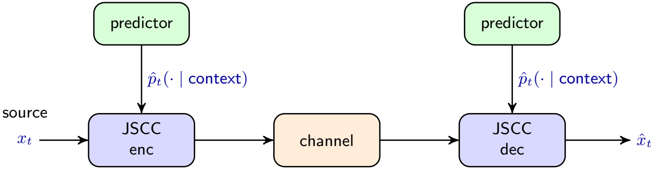


- Fixed-slope universal lossy data compression [Yang, Zhang, Berger '97]

$$\hat{x}^n = \arg\min_{\tilde{x}^n} \mathrm{LZ}(\tilde{x}^n) + s d(x^n, \tilde{x}^n)$$

- Universal zero-delay JSCC [Matloub & Weissman '06]: regret analysis
- JSCC at finite blocklength [Kostina & Verdú, '13]: gain over separate schemes

Related Problems



- Fixed-slope universal lossy data compression [Yang, Zhang, Berger '97]

$$\hat{x}^n = \arg\min_{\tilde{x}^n} \mathrm{LZ}(\tilde{x}^n) + s d(x^n, \tilde{x}^n)$$

- Universal zero-delay JSCC [Matloub & Weissman '06]: regret analysis
- JSCC at finite blocklength [Kostina & Verdú, '13]: gain over separate schemes
- Online (adaptive) conformal inference [Gibbs & Candes '24; Feldman et al., '23]: prediction sets with coverage guarantees for sequential data

This Talk: Two Special Cases

- Noiseless channel

This Talk: Two Special Cases

- Noiseless channel

  - Outage distortion $d(x, \hat{x}) = \mathbb{1}\{\hat{x} \neq x\}$

This Talk: Two Special Cases

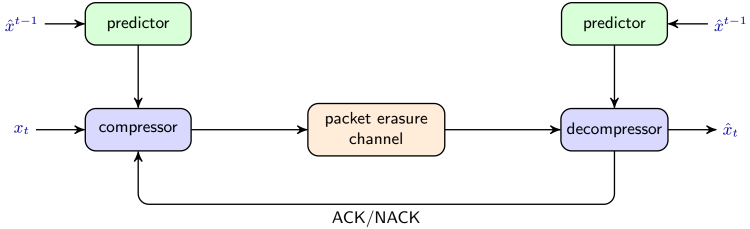
- Noiseless channel

  - Outage distortion $d(x, \hat{x}) = \mathbb{1}\{\hat{x} \neq x\}$
  - Arbitrary bounded distortion $d(x, \hat{x}) \in [0, D_{\max}]$

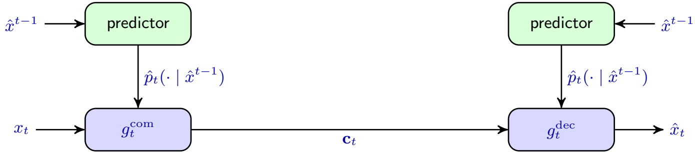
This Talk: Two Special Cases

- Noiseless channel

  - Outage distortion $d(x,\hat{x}) = \mathbb{1}\{\hat{x} \neq x\}$
  - Arbitrary bounded distortion $d(x,\hat{x}) \in [0, D_{\max}]$
- Packet erasure channel with perfect ACK/NACK feedback

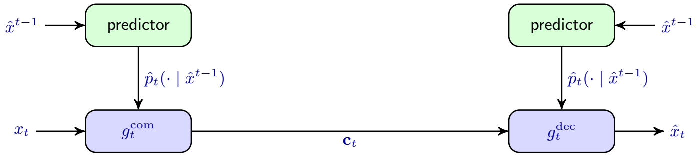


Noiseless Channel with Outage Distortion: Setup



- Variable-length compressor: $g_t^{\text{com}} : \mathcal{X} \times \mathcal{P}(\mathcal{X}) \to \{0,1\}^*$; $\quad g_t^{\text{com}}(x_t, p_t(\cdot \,|\, \hat{x}^{t-1})) = \mathbf{c}_t$
- Decompressor: $g_t^{\text{dec}} : \{0,1\}^* \times \mathcal{P}(\mathcal{X}) \to \mathcal{X}$; $\quad g_t^{\text{dec}}(\mathbf{c}_t, p_t(\cdot \,|\, \hat{x}^{t-1})) = \hat{x}_t$
- Per-symbol outage distortion: $d(x_t, \hat{x}_t) = \mathbb{1}\{\hat{x}_t \neq x_t\}$

Noiseless Channel with Outage Distortion: Goal

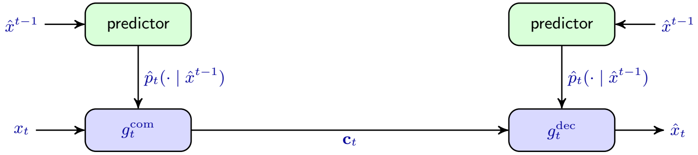


Goal

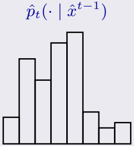
- Minimize compression rate $R_n = \dfrac{1}{n}\sum_{t=1}^{n} \ell(\mathbf{c}_t)$
- Subject to distortion constraint

$$\frac{1}{n}\sum_{t=1}^{n} \mathbb{1}\{\hat{x}_t \neq x_t\} \leq D + \frac{\mathrm{c}}{n}, \quad \forall n,\ \forall x^n,\ \forall \{\hat{p}_t(\cdot \mid \cdot)\}_{t=1}^{n}$$

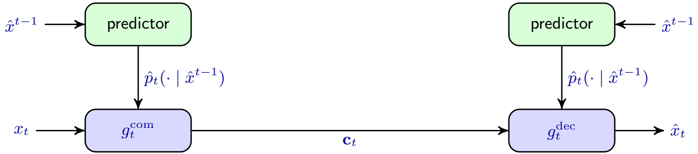
An Adaptive Conformal Inference Approach



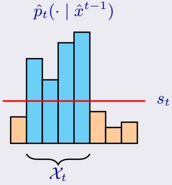
Compressor operations



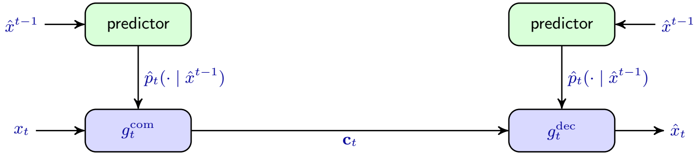
An Adaptive Conformal Inference Approach



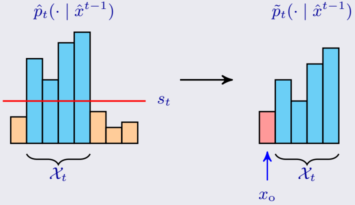
Compressor operations



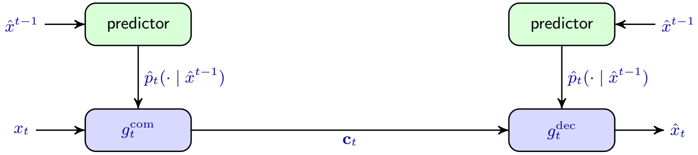
An Adaptive Conformal Inference Approach



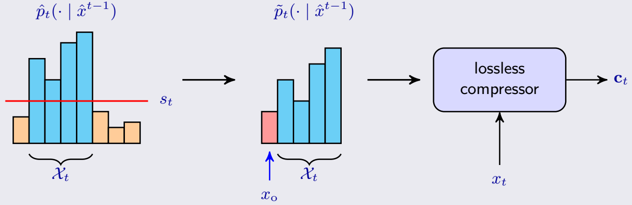
Compressor operations



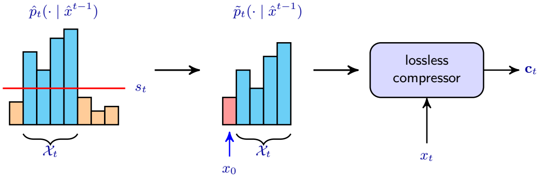
An Adaptive Conformal Inference Approach



Compressor operations

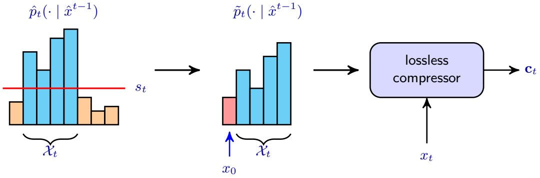


An Adaptive Conformal Inference Approach: Compressor



- Define high-probability set $\mathcal{X}_t = \{x \in \mathcal{X} : \hat{p}_t(x \,|\, \hat{x}^{t-1}) > s_t\}$

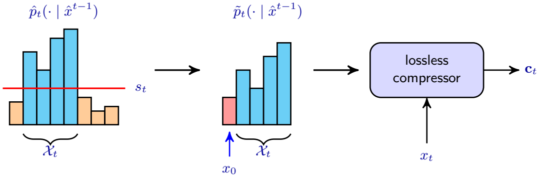
An Adaptive Conformal Inference Approach: Compressor



- Define high-probability set $\mathcal{X}_t = \{x \in \mathcal{X} : \hat{p}_t(x \mid \hat{x}^{t-1}) > s_t\}$
- Renormalize probability distribution after adding outage symbol x_{o} to $\mathcal{X}_t$

$$\tilde{p}_t(x \mid \hat{x}^{t-1}) = \begin{cases} (1-D)\dfrac{\hat{p}_t(x \mid \hat{x}^{t-1})}{\sum_{x' \in \mathcal{X}_t} \hat{p}_t(x \mid \hat{x}^{t-1})}, & x \in \mathcal{X}_t \\ D, & x = x_{\mathrm{o}} \end{cases}$$

An Adaptive Conformal Inference Approach: Compressor

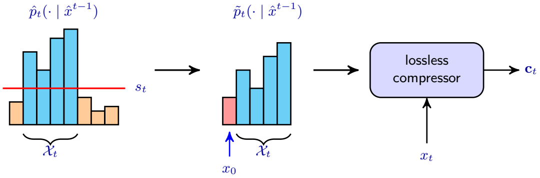


- Define high-probability set $\mathcal{X}_t = \{x \in \mathcal{X} : \hat{p}_t(x \mid \hat{x}^{t-1}) > s_t\}$
- Renormalize probability distribution after adding outage symbol x_o to $\mathcal{X}_t$

$$\tilde{p}_t(x \mid \hat{x}^{t-1}) = \begin{cases} (1-D)\dfrac{\hat{p}_t(x \mid \hat{x}^{t-1})}{\sum_{x' \in \mathcal{X}_t} \hat{p}_t(x \mid \hat{x}^{t-1})}, & x \in \mathcal{X}_t \\ D, & x = x_\mathrm{o} \end{cases}$$

- Construct lossless, prefix-free code $\{\mathbf{c}(x), x \in \mathcal{X}_t \cup \{x_\mathrm{o}\}\}$ matched to $\tilde{p}(\cdot \mid \hat{x}^{t-1})$

An Adaptive Conformal Inference Approach: Compressor

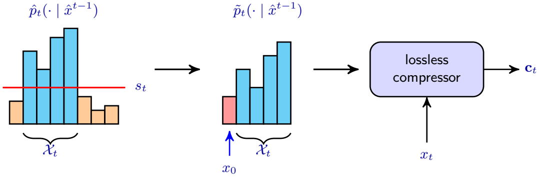


- Define high-probability set $\mathcal{X}_t = \{x \in \mathcal{X} : \hat{p}_t(x \,|\, \hat{x}^{t-1}) > s_t\}$
- Renormalize probability distribution after adding outage symbol x_o to $\mathcal{X}_t$

$$\tilde{p}_t(x \,|\, \hat{x}^{t-1}) = \begin{cases} (1-D)\dfrac{\hat{p}_t(x \,|\, \hat{x}^{t-1})}{\sum_{x' \in \mathcal{X}_t} \hat{p}_t(x \,|\, \hat{x}^{t-1})}, & x \in \mathcal{X}_t \\ D, & x = x_\mathrm{o} \end{cases}$$

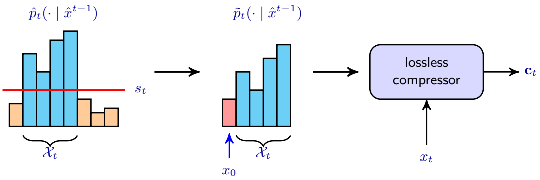
- Construct lossless, prefix-free code $\{\mathbf{c}(x), x \in \mathcal{X}_t \cup \{x_\mathrm{o}\}\}$ matched to $\tilde{p}(\cdot \,|\, \hat{x}^{t-1})$
- Set $\mathbf{c}_t = \mathbf{c}(x_t)$ if $x_t \in \mathcal{X}_t$ and $\mathbf{c}_t = \mathbf{c}(x_\mathrm{o})$ otherwise

Decompressor and Threshold Update



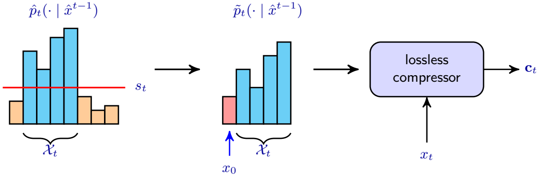
- Given s_t and $\hat{p}(\cdot \,|\, \hat{x}^{t-1})$, construct $\mathcal{X}_t$, $\tilde{p}(x \,|\, \hat{x}^{t-1})$, and code $\{\mathbf{c}(x), x \in \mathcal{X}_t \cup \{x_o\}\}$

Decompressor and Threshold Update



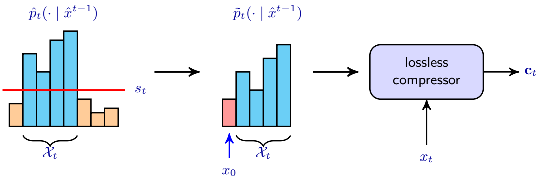
- Given s_t and $\hat{p}(\cdot \,|\, \hat{x}^{t-1})$, construct $\mathcal{X}_t$, $\tilde{p}(x \,|\, \hat{x}^{t-1})$, and code $\{\mathbf{c}(x), x \in \mathcal{X}_t \cup \{x_\text{o}\}\}$
- If $x_t \in \mathcal{X}$ is the symbol associated to $\mathbf{c}_t$, set $\hat{x}_t = x_t$

Decompressor and Threshold Update



- Given s_t and $\hat{p}(\cdot \,|\, \hat{x}^{t-1})$, construct $\mathcal{X}_t$, $\tilde{p}(x \,|\, \hat{x}^{t-1})$, and code $\{\mathbf{c}(x), x \in \mathcal{X}_t \cup \{x_\mathrm{o}\}\}$
- If $x_t \in \mathcal{X}$ is the symbol associated to $\mathbf{c}_t$, set $\hat{x}_t = x_t$
- If x_o is the symbol associated to $\mathbf{c}_t$, set $\hat{x}_t = \arg\max_{x \notin \mathcal{X}_t} \hat{p}_t(x \,|\, \hat{x}^{t-1})$

Decompressor and Threshold Update



- Given s_t and $\hat{p}(\cdot \,|\, \hat{x}^{t-1})$, construct $\mathcal{X}_t$, $\tilde{p}(x \,|\, \hat{x}^{t-1})$, and code $\{\mathbf{c}(x), x \in \mathcal{X}_t \cup \{x_\mathrm{o}\}\}$
- If $x_t \in \mathcal{X}$ is the symbol associated to $\mathbf{c}_t$, set $\hat{x}_t = x_t$
- If x_o is the symbol associated to $\mathbf{c}_t$, set $\hat{x}_t = \arg\max_{x \notin \mathcal{X}_t} \hat{p}_t(x \,|\, \hat{x}^{t-1})$
- Adaptive conformal prediction threshold update at both compressor and decompressor

$$s_t = \max\{0, \lambda_t\}$$

$$\lambda_{t+1} = \lambda_t - \underset{\text{step size}}{\eta}\,(\mathbb{1}\{x_t \notin \mathcal{X}_t\} - D)$$

Theoretical Guarantees

Theorem

For every sequence of source symbols x^n and predictors $\{\hat{p}_t\}_{t=1}^n$ and every target distortion $D \in [0,1]$, the distortion of the reconstructed sequence $\hat{x}^n$ satisfies

positive initialization

$$\frac{1}{n}\sum_{t=1}^{n} \mathbb{1}\{x_t \neq \hat{x}_t\} \leq D + \frac{\eta(1-D) + \lambda_0}{\eta n}$$

Theoretical Guarantees

Theorem

For every sequence of source symbols x^n and predictors $\{\hat{p}_t\}_{t=1}^n$ and every target distortion $D \in [0,1]$, the distortion of the reconstructed sequence $\hat{x}^n$ satisfies

$$\frac{1}{n}\sum_{t=1}^{n} \mathbb{1}\{x_t \neq \hat{x}_t\} \leq D + \frac{\eta(1-D) + \overset{\substack{\text{positive initialization}\\ \downarrow}}{\lambda_0}}{\eta n}$$

Proof

- Recall the threshold update rule

$$\lambda_{t+1} = \lambda_t - \eta\left(\mathbb{1}\{x_t \notin \mathcal{X}_t\} - D\right)$$

- Telescoping

$$\frac{1}{n}\sum_{t=1}^{n} \mathbb{1}\{x_t \neq \hat{x}_t\} \leq D + \frac{\lambda_0 - \lambda_{n+1}}{\eta n}$$

- Show that λ_t is bounded from below uniformly in t

Proof of last statement

Claim: $\lambda_t \geq -\eta(1-D)$ for all t

- Assume that the sequence $\{\lambda_t\}$ drops below $-\eta(1-D)$ for the first time at $t'+1$

Proof of last statement

Claim: $\lambda_t \geq -\eta(1-D)$ for all t

- Assume that the sequence $\{\lambda_t\}$ drops below $-\eta(1-D)$ for the first time at $t'+1$
- Then $\lambda_{t'} = \lambda_{t'+1} + \eta\left(\mathbb{1}\{x_t \notin \mathcal{X}_t\} - D\right) < 0$

Proof of last statement

Claim: $\lambda_t \geq -\eta(1-D)$ for all t

- Assume that the sequence $\{\lambda_t\}$ drops below $-\eta(1-D)$ for the first time at $t'+1$
- Then $\lambda_{t'} = \lambda_{t'+1} + \eta\,(\mathbb{1}\{x_t \notin \mathcal{X}_t\} - D) < 0$
- Hence, $s_{t'} = \max\{0, \lambda_{t'}\} = 0$ and $\mathcal{X}_{t'} = \mathcal{X}$

Proof of last statement

Claim: $\lambda_t \geq -\eta(1-D)$ for all t

- Assume that the sequence $\{\lambda_t\}$ drops below $-\eta(1-D)$ for the first time at $t'+1$
- Then $\lambda_{t'} = \lambda_{t'+1} + \eta\left(\mathbb{1}\{x_t \notin \mathcal{X}_t\} - D\right) < 0$
- Hence, $s_{t'} = \max\{0, \lambda_{t'}\} = 0$ and $\mathcal{X}_{t'} = \mathcal{X}$
- As a consequence, $\lambda_{t'+1} = \lambda_{t'} - \eta\left(\mathbb{1}\{x_t \notin \mathcal{X}_t\} - D\right) = \lambda_{t'} + \eta D$

Proof of last statement

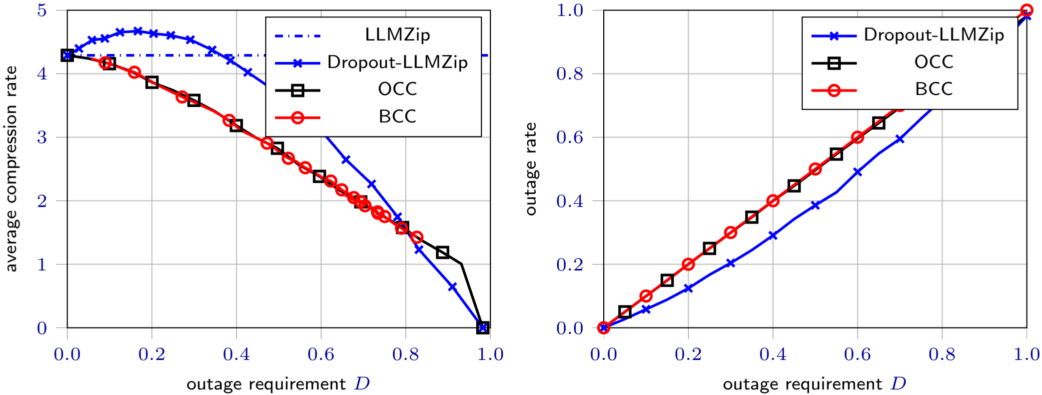
Claim: $\lambda_t \geq -\eta(1-D)$ for all t

- Assume that the sequence $\{\lambda_t\}$ drops below $-\eta(1-D)$ for the first time at $t'+1$
- Then $\lambda_{t'} = \lambda_{t'+1} + \eta\left(\mathbb{1}\{x_t \notin \mathcal{X}_t\} - D\right) < 0$
- Hence, $s_{t'} = \max\{0, \lambda_{t'}\} = 0$ and $\mathcal{X}_{t'} = \mathcal{X}$
- As a consequence, $\lambda_{t'+1} = \lambda_{t'} - \eta\left(\mathbb{1}\{x_t \notin \mathcal{X}_t\} - D\right) = \lambda_{t'} + \eta D$
- Contradiction!

Some Numerical Results: Online Conformal Compression (OCC)

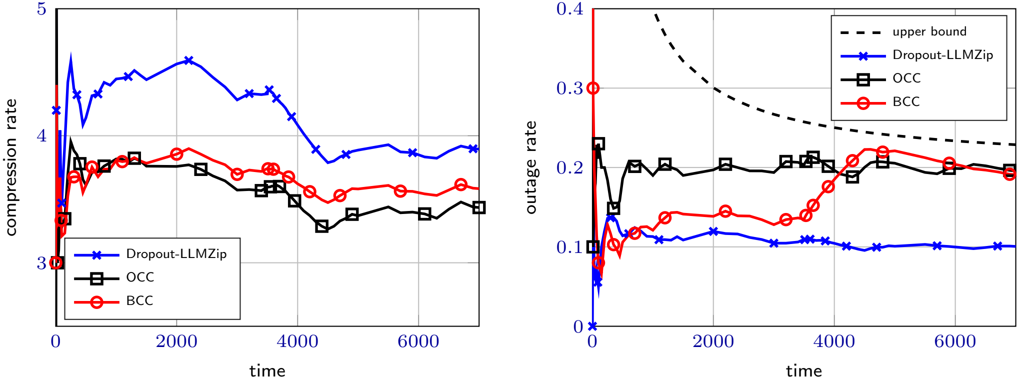
- Lossy compression on English text
- Predictor: nanoGPT-2 model fine-tuned on holdout portion of Shakespeare's work
- $|\mathcal{X}| = 50\,257$ tokens
- Baseline
  - Dropout-LLMZip: lossless compression $+$ symbols dropped independently with probability D
  - Block-conformal compression (BCC): chooses optimal threshold s^* in hindsight, based on entire sequence x^n

Performance on Shakespeare's Dataset

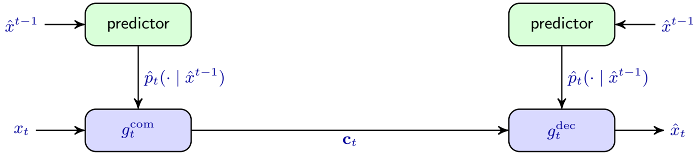


Performance on Nonstationary Data

Up to time $n = 3500$ data comes from Shakespeare's dataset and afterwards from Taylor Swift's songs



More General Distortion Metrics

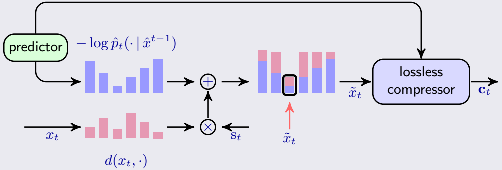


- The online conformal compression scheme just described relies heavily on the binary nature of the outage distortion
- Both compressor and decompressor can track accumulated distortion!
- For more general distortion functions this is possible only at the compressor side
- Key idea: combine (per-symbol) fixed-slope universal lossy compression metric with adaptive conformal inference

A Variable-Slope Scheme

Compressor

- Input: slope parameter s_t, variable-length prefix-free code $\{\mathbf{c}(x), x \in \mathcal{X}\}$ matched to $\hat{p}(\cdot \,|\, \hat{x}^{t-1})$



- Compute

$$\tilde{x}_t = \arg\min_{x \in \mathcal{X}} \left\{ -\log \hat{p}_t(x \,|\, \hat{x}^{t-1}) + s_t d(x_t, x) \right\}$$

- Set $\mathbf{c}_t = \mathbf{c}(\tilde{x}_t)$

A Variable-Slope Scheme

Compressor

$$\tilde{x}_t = \arg\min_{x \in \mathcal{X}} \left\{ -\log \hat{p}_t(x \mid \hat{x}^{t-1}) + s_t d(x_t, x) \right\}$$

Decompressor

- Construct variable length prefix-free code $\{\mathbf{c}(x), x \in \mathcal{X}\}$ matched to $\hat{p}(\cdot \mid \hat{x}^{t-1})$
- Invert $\mathbf{c}_t$ to recover $\hat{x}_t = \tilde{x}_t$

Threshold update (at compressor only)

$$s_t = \max\{0, \lambda_t\}$$
$$\lambda_{t+1} = \lambda_t + \eta\big(d(x_t, \hat{x}_t) - D\big)$$

Theoretical Guarantees

Theorem

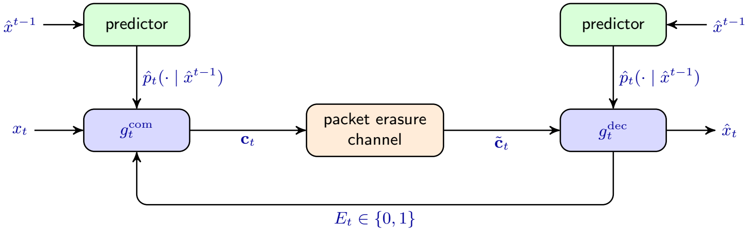
Assume that for all t and all $\hat{x}^{t-1}$, there exists a constant L such that

$$\max_{x \in \mathcal{X}} -\log \hat{p}_t(x \,|\, \hat{x}^{t-1}) < L$$

Then for every x^n, every sequence of predictors satisfying the assumption, and every target distortion level $D \in (0, D_{\max}]$, the distortion of the reconstructed sequence satisfies

$$\frac{1}{n}\sum_{t=1}^{n} d(x_t, \hat{x}_t) \le D + \frac{L/D + \eta(D_{\max} - D) - \lambda_0}{\eta n}$$

Packet Erasure Channel



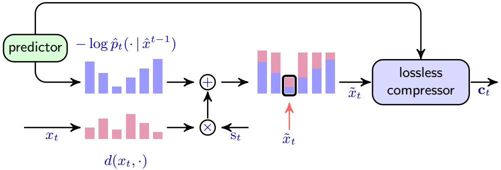
- Erasure model: $\tilde{\mathbf{c}}_t = \mathbf{c}_t$ if $E_t = 0$, and $\tilde{\mathbf{c}}_t = \circledast$ if $E_t = 1$
- Decoder: if $\tilde{\mathbf{c}}_t = \circledast$, set $\hat{x}_t = \arg\max_{x \in \mathcal{X}} \hat{p}_t(x \,|\, \hat{x}^{t-1})$
- Channel induced distortion known to encoder via ACK/NACK feedback

$$\delta_t = d(x_t, \hat{x}_t) - d(x_t, \tilde{x}_t) \geq 0$$

with

$$\tilde{x}_t = \arg\min_{x \in \mathcal{X}} \big\{ -\log \hat{p}_t(x \,|\, \hat{x}^{t-1}) + s_t d(x_t, x) \big\}$$

A Doubly-Adaptive Conformal Update Rule



- Accounting for channel-induced distortion: fix $\epsilon > 0$ and let $s_t = \max\{0, \lambda_t\}$ with

correction

$$\lambda_{t+1} = \lambda_t + \eta(d(x_t, \tilde{x}_t) - (D - \delta_t^{\text{cor}}))$$

$$\delta_{t+1}^{\text{cor}} = \min\{D - \epsilon, \Delta_t^{\text{ch}} - \Delta_t^{\text{cor}}\}$$

$$\Delta_t^{\text{ch}} = \sum_{i=1}^{t} \delta_t^{\text{ch}}; \quad \Delta_t^{\text{cor}} = \sum_{i=1}^{t} \delta_t^{\text{cor}};$$

- Queuing analogy: drain residual distortion $Q_t = \Delta_t^{\text{ch}} - \Delta_t^{\text{cor}}$ under maximum service rate $D - \epsilon$

Theoretical Guarantees

Theorem

For every sequence of symbols x^n and every sequence of predictors $\hat{p}_t$ satisfying the boundness assumption, and for every target distortion level $D \in (0, D_{\max}]$, the average distortion of the reconstructed sequence $\hat{x}_t$ satisfies

$$\frac{1}{n}\sum_{t=1}^{n} d(x_t, \hat{x}_t) \leq D + \frac{L/\epsilon + \eta(D_{\max} - \epsilon) - \lambda_0}{\eta n} + \frac{Q_n}{n}$$

Question

Under which conditions does Q_n grow sublinearly in n?

Minimum Envelope Process

Definition

Given the sequence of erasure events $\{E_t\}_{t\geq 1}$ and an integer $\tau \geq 1$, the minimum envelope process is defined as the maximum number of erasures within any window of τ consecutive slots

$$\Psi(\tau, \{E_t\}) = \sup_{k\geq 1} \sum_{t=k}^{k+\tau} E_t$$

Minimum Envelope Process

Definition

Given the sequence of erasure events $\{E_t\}_{t\geq 1}$ and an integer $\tau \geq 1$, the minimum envelope process is defined as the maximum number of erasures within any window of τ consecutive slots

$$\Psi(\tau, \{E_t\}) = \sup_{k\geq 1} \sum_{t=k}^{k+\tau} E_t$$

Sufficient condition for sublinear growth of Q_n

There exists a constant $A < (D - \epsilon)/D_{\max}$ and a sublinear function $\phi(\tau)$ such that

$$\Psi(\tau, \{E_t\}) < A\tau + \phi(\tau)$$

Theoretical Guarantees via Minimum Envelope Process

Sufficient condition for sublinear growth of Q_n

There exists a constant $A < (D - \epsilon)/D_{\max}$ and a sublinear function $\phi(\tau)$ such that

$$\Psi(\tau, \{E_t\}) < A\tau + \phi(\tau)$$

Theorem

Suppose the sequence of the erasure events $\{E_t\}$ satisfy the sufficient condition just stated; then for every sequence of source symbols x^n and predictors $\{\hat{p}_t\}$ and target distortion level $D \in (0, D_{\max}]$, we have that

$$Q_n \leq \tau_{\max}(D_{\max} - D - \epsilon) + D - \epsilon$$

where

$$\tau_{\max} = \min\left\{\tau \in \mathbb{N} : \frac{\phi(\tau)}{\tau} \leq \frac{D - \epsilon}{D_{\max}} - A\right\}$$

Proof

- Consider w.l.o.g. an interval $[t_1, t_2)$ of size τ in which $Q_t > D - \epsilon$ for all $t \in [t_1, t_2)$; $Q_{t_1-1} < D - \epsilon$, $Q_{t_2} < D - \epsilon$; we want to bound τ

Proof

- Consider w.l.o.g. an interval $[t_1, t_2)$ of size τ in which $Q_t > D - \epsilon$ for all $t \in [t_1, t_2)$; $Q_{t_1 - 1} < D - \epsilon$, $Q_{t_2} < D - \epsilon$; we want to bound τ
- Note that

$$Q_{t_2} = \sum_{t=t_1}^{t_2} (\delta_t^{\mathrm{ch}} - \delta_t^{\mathrm{cor}})$$

Proof

- Consider w.l.o.g. an interval $[t_1, t_2)$ of size τ in which $Q_t > D - \epsilon$ for all $t \in [t_1, t_2)$; $Q_{t_1 - 1} < D - \epsilon$, $Q_{t_2} < D - \epsilon$; we want to bound τ
- Note that

$$Q_{t_2} = \sum_{t=t_1}^{t_2} \delta_t^{\mathrm{ch}} - (\tau - 1)(D - \epsilon)$$

Proof

- Consider w.l.o.g. an interval $[t_1, t_2)$ of size τ in which $Q_t > D - \epsilon$ for all $t \in [t_1, t_2)$; $Q_{t_1-1} < D - \epsilon$, $Q_{t_2} < D - \epsilon$; we want to bound τ
- Note that

$$Q_{t_2} \leq D_{\max} \sum_{t=t_1}^{t_2} E_t - (\tau - 1)(D - \epsilon)$$

Proof

- Consider w.l.o.g. an interval $[t_1, t_2)$ of size τ in which $Q_t > D - \epsilon$ for all $t \in [t_1, t_2)$; $Q_{t_1 - 1} < D - \epsilon$, $Q_{t_2} < D - \epsilon$; we want to bound τ
- Note that

$$Q_{t_2} \leq D_{\max}[A\tau + \phi(\tau)] - (\tau - 1)(D - \epsilon)$$

Proof

- Consider w.l.o.g. an interval $[t_1, t_2)$ of size τ in which $Q_t > D - \epsilon$ for all $t \in [t_1, t_2)$; $Q_{t_1 - 1} < D - \epsilon$, $Q_{t_2} < D - \epsilon$; we want to bound τ
- Note that

$$Q_{t_2} \leq (AD_{\max} - D + \epsilon)\tau + D_{\max}\phi(\tau) + D - \epsilon$$

Proof

- Consider w.l.o.g. an interval $[t_1, t_2)$ of size τ in which $Q_t > D - \epsilon$ for all $t \in [t_1, t_2)$; $Q_{t_1 - 1} < D - \epsilon$, $Q_{t_2} < D - \epsilon$; we want to bound τ
- Note that

$$Q_{t_2} \leq (AD_{\max} - D + \epsilon)\tau + D_{\max}\phi(\tau) + D - \epsilon$$

- Drops below $D - \epsilon$ at

$$\tau_{\max} = \min\left\{\tau \in \mathbb{N} : \frac{\phi(\tau)}{\tau} \leq \frac{D - \epsilon}{D_{\max}} - A\right\}$$

Random Erasures and Guarantees via Concentration of Measure

Memoryless time-varying erasure channel

- Assume that $\{E_t \sim \text{Bern}(e_t)\}$, that the $\{E_t\}$ are independent and that the $\{e_t\}$ satisfy minimum envelope assumption
$$\Psi(\tau, \{e_t\}) < A\tau + \phi(\tau)$$
with $A < (D - \epsilon)/D_{\max}$ and $\phi(\tau)$ sublinear.
- Then for every sequence of source symbols x^n, every sequence of predictors $\{\hat{p}_t\}$, and every distortion level $D \in (0, D_{\max}]$, for every $\delta_n \in (0, 1)$ the distortion of the reconstructed sequence $\hat{x}^n$ satisfies
$$\frac{1}{n}\sum_{t=1}^{n} d(x_t, \hat{x}_t) \leq D + \frac{L/\epsilon + \eta(D_{\max} - \epsilon) - \lambda_0}{\eta n} + \mathcal{O}\left(\frac{D_{\max}\phi(n)}{n} + \frac{D_{\max}}{n}\log(1/\delta_n)\right)$$
with probability $1 - \delta_n$

Erasures in Bursts: Gilbert-Elliot Erasure Model

- Two-state Markov chain with state space $\{B, G\}$; e_{B}: erasure prob. corresponding to B; e_{G}: erasure prob. corresponding to B
- $[\pi(B), \pi(G)]$: steady state probability; $\bar{e} = \pi(B)e_{\mathrm{B}} + \pi(G)e_{\mathrm{G}}$: steady-state erasure prob.
- γ: spectral gap

Lemma: concentration of measure for minimum envelope of $\{e_t\}$

Assume that $\gamma < 1$; for every $\delta \in (0,1]$ for all $\tau > 1$ and all $k \in \{1, \dots, n\}$,

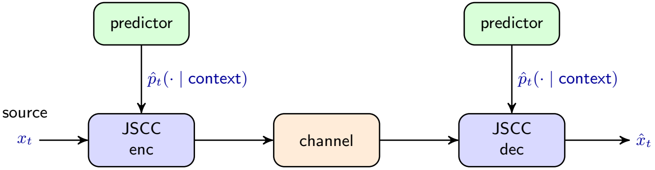
$$\sum_{t=k}^{k+\tau} e_t \leq \bar{e}\tau + \sqrt{\tau \frac{1+\gamma}{1-\gamma} \log\left(\frac{\sqrt{n}\pi^2\tau^2}{6\delta \min\{\pi(B), \pi(G)\}}\right)}$$

with probability $1-\delta$

Implication

Probabilistic distortion guarantees provided that $\bar{e} < (D-\epsilon)/D_{\max}$

Conclusion



Summary

- Prediction-aided communication (noiseless and packet erasure)
- Sequential, universal distortion guarantees
- Novel schemes based on adaptive conformal inference

Open questions

- More realistic channel models and JSCC
- Multiple predictors
- Online predictor adaptation and regret analysis
- Efficient implementations