Single-Shot Compression for Hypothesis Testing

Fabrizio Carpi, Siddharth Garg, Elza Erkip

Introduction

Motivation: a resource constrained client offloads costly task-related computations to remote a server (edge/cloud computing).

Open need: design task-aware source coding schemes which provides effective representations of the source data.

Assumptions:
- task: binary hypothesis testing;
- client: constrained device which cannot perform task locally, does not have memory and can only do simple scalar compression;
- server: hypothesis testing on a block of compressed samples.

Our work: single-shot fixed-length compression for hypothesis testing.

- problem formulation;
- analyze the error performance;
- propose a task-oriented compression algorithm for hypothesis testing.

System Model

Hypothesis Testing under Single-shot Compression

Hypothesis test on $X \sim \hat{P}_0$:
- log-likelihood ratio test on $\hat{X}^*$ is optimal;
- optimal error exponent is $\gamma^*(R) = D(\hat{P}_0 || \hat{P}_1)$.

Compression penalty: $\Delta f(R) = D(\hat{P}_0 || P_1) - D(\hat{P}_0 || \hat{P}_1)$

Proposition 1. Expression for $\Delta f \geq 0$:

$$\Delta f = \sum_{x=1}^{M} \hat{P}_1(x) D\left(\hat{P}_0(x) || P_1(x | \hat{X})\right)$$

$P_1(X | \hat{X}) = \frac{P_1(X)}{P_1(\hat{X})}$ is the posterior of $X$ given $\hat{X} = i(X)$.

Note that a good task-aware compression strategy combines $X$ that have similar posteriors $P_0(X | \hat{X})$.

Optimal compressor:

$\hat{f}^* = \arg\max_i D(\hat{P}_0 || \hat{P}_i) = \arg\min_i \Delta f$ s.t. $|i| \leq M$;

optimization over each possible $i$, which induces a partition of $M$ sets over $X$ (NP-hard).

Proposed Compressor Scheme

Optimal one-step compression from $|X|$ to $|X| - 1$:
- $i$ combines $\{a, b\} \subset X$ and the others in $X \setminus \{a, b\}$ are one-to-one;
- $i.e.$, $i(a) = i(b) = m \in M, \text{if } i = i \in M \setminus \{m\}$;

Thus,

$$\hat{f}^* = \arg\min_{\{a,b\} \subset X \setminus \{i\}} \left\{ \hat{P}_0(m) D\left(\hat{P}_0(x | m) || P_1(x | m)\right) \right\}$$

Our "KL-greedy" compressor:
- iteratively reduce the alphabet size by 1 at each step, until the compressed alphabet has size $M$;
- at each step, combine $\{a, b\}$ which minimize (1);
- note that this compressor can be determined in polynomial time.

Results

$P_i$ are shifted binomial distributions with different parameters.

Compare compression penalty $\Delta f$ and empirical type-II error rate for:
- optimal compressor $\hat{f}^*$ — when feasible to compute, i.e, small $|X|$;
- our KL-greedy compressor;
- universal compressor from [2], which is designed for reconstruction under log-loss distortion.

For the empirical type-II error rate, consider equal priors and $T = 1$.

References


