Lossy Compression of Permutations

Da Wang EECS, MIT

Arya Mazumdar ECE, Univ. Minnesota

Gregory W. Wornell EECS, MIT

ISIT 2014, Honolulu, HI

June 30, 2014

Outline

1. Why lossy compression of permutations
   - Storage of ranking data
   - Analysis of approximate sorting algorithms
2. Rate distortion problem in permutation space
   - Worst-case and average-case
   - Distortion measures of interest
3. Results
   - Relationship among distortion measures
   - Equivalence among source codes
   - Lossy compression schemes

Storage of ranking data

- Permutation $\sigma = [3, 4, 1, 2, 5]$

Ranking as a permutation σ

- A list of items $v_1, v_2, \ldots, v_n$ such that

$$v_{\sigma^{-1}(1)} \succ v_{\sigma^{-1}(2)} \succ \ldots \succ v_{\sigma^{-1}(n)}$$

- σ: the ranking of these list of items
  - $\sigma(i)$: rank of item v_i
  - $\sigma^{-1}(r)$: the index of the item with rank r

Recommendation systems

- Storing preferences of all users
- Rough knowledge may be sufficient

Storage of ranking data

- Permutation $\sigma = [3, 4, 1, 2, 5]$

Ranking as a permutation σ

- A list of items $v_1, v_2, \ldots, v_n$ such that

$$v_{\sigma^{-1}(1)} \succ v_{\sigma^{-1}(2)} \succ \ldots \succ v_{\sigma^{-1}(n)}$$

- σ: the ranking of these list of items
  - $\sigma(i)$: rank of item v_i
  - $\sigma^{-1}(r)$: the index of the item with rank r

Recommendation systems

- Storing preferences of all users
- Rough knowledge may be sufficient

lossy compression!

Given a certain distortion measure,

Lossy compression:

- need R bits to describe σ up to distortion D

Approximate sorting:

- Assume: all elements are distinct
- Comparison-based sorting: search for the "true" ordering (permutation)
- A comparison provides at most 1 bit of information
- Need at least R comparisons to find a permutation with distortion D

Analysis of approximate sorting algorithms

Given a certain distortion measure,

Lossy compression:

- need R bits to describe σ up to distortion D

Approximate sorting:

- Assume: all elements are distinct
- Comparison-based sorting: search for the "true" ordering (permutation)
- A comparison provides at most 1 bit of information
- Need at least R comparisons to find a permutation with distortion D

An information-theoretic lower bound on query complexity

Rate-distortion theory of a permutation space

- First formulated in [**W.**, Mazumdar & Wornell, ISIT'13]

Permutation space

- $\mathcal{S}_n$: the set of $n!$ permutations
- d: distance measure

(n, D_n) source code $\mathcal{C}_n$

- $\mathcal{C}_n \subset \mathcal{S}_n$
- Encoder: $f_n : \mathcal{S}_n \to \mathcal{C}_n$

Worst-case distortion:

$$\max_{\sigma} d(\sigma, f_n(\sigma)) \leq D_n.$$

Average-case distortion:

$$\mathbb{E}\left[d(\sigma, f_n(\sigma))\right] \leq D_n.$$

- Assume uniform distribution over $\mathcal{S}_n$

Rate-distortion function

Let $A(n, D_n)$ be the minimum size of the (n, D_n) source codes with distortion D_n. The minimal rate for distortion D_n is

$$R(D_n) \triangleq \frac{\log A(n, D_n)}{\log n!},$$

- Under average-case distortion: $\bar{R}(D_n)$
- Under worst-case distortion: $\hat{R}(D_n)$

Four distance measures of interest

Among the many possibilities…

1. ℓ_∞ distance of permutation vectors (Chebyshev distance)
   - Maximum of rank deviations
2. ℓ_1 distance of permutation vectors (Spearman's footrule)
   - Sum of rank deviations
3. Kendall tau distance of permutation vectors
   - Number of "operations" to eliminate rank deviations
   - [**W.**, Mazumdar & Wornell, ISIT'13]
4. ℓ_1 distance of inversion vectors (inversion-ℓ_1 distance)
   - Inversion vector: keeps track of "out-of-order" elements in the permutation
   - [**W.**, Mazumdar & Wornell, ISIT'13]

Results in this talk

- After scaling, these distortion measures lower and upper bound each other
  - Sometimes in a probabilistic sense
- Lead to
  - equivalence between source codes
  - similar rate-distortion functions
- Lossy compression schemes

Given two permutations σ_1 and σ_2,

$$d_{\ell_\infty}(\sigma_1, \sigma_2) \triangleq \|\sigma_1 - \sigma_2\|_\infty = \max_{1 \le i \le n} |\sigma_1(i) - \sigma_2(i)|$$

$$d_{\ell_1}(\sigma_1, \sigma_2) \triangleq \|\sigma_1 - \sigma_2\|_1 = \sum_{i=1}^{n} |\sigma_1(i) - \sigma_2(i)|$$

- The *Kendall tau distance* $d_\tau(\sigma_1, \sigma_2)$:
  the minimum number of swaps of adjacent elements required to change σ_1 into σ_2.

Properties

- upper bounded by $\binom{n}{2}$
- $d_\tau\,(\sigma, e)$ = number of swaps in bubble sort

Inversion

- An *inversion* in a permutation σ: a pair $(\sigma(i), \sigma(j))$ such that $i < j$ and $\sigma(i) > \sigma(j)$.
  - Inversions in $\sigma_1 = [1, 5, 4, 2, 3]$: $(5, 4), (5, 2), (5, 3), (4, 2), (4, 3)$
  - Inversions in $\sigma_2 = [3, 4, 5, 1, 2]$: $(3, 1), (3, 2), (4, 1), (4, 2), (5, 1), (5, 2)$

Inversion vector $\mathbf{x}_\sigma \in [0:1] \times [0:2] \times \cdots \times [0:n-1]$

$\mathbf{x}_\sigma(i) =$ the number of inversions in σ in which $i + 1$ is the first element

$$i = 1, 2, \ldots, n-1.$$

Examples

$$\sigma_1 = [1, 5, 4, 2, 3] \Rightarrow \mathbf{x}_{\sigma_1} = [0, 0, 2, 3]$$
$$\sigma_2 = [3, 4, 5, 1, 2] \Rightarrow \mathbf{x}_{\sigma_2} = [0, 2, 2, 2]$$
$$d_{\mathbf{x}, \ell_1}(\sigma_1, \sigma_2) = d_{\ell_1}([0, 0, 2, 3], [0, 2, 2, 2]) = 3$$

- Inversion vector: a common measure of sortedness

Relationship between distortion measures

For any two permutations σ_1 and σ_2 in $\mathcal{S}_n$,

$$n \cdot d_{\ell_\infty}(\sigma_1, \sigma_2) \;\geq\; d_{\ell_1}(\sigma_1, \sigma_2) \;\overset{(a)}{\geq}\; d_\tau\left(\sigma_1^{-1}, \sigma_2^{-1}\right) \;\geq\; d_{\mathbf{x},\ell_1}\left(\sigma_1^{-1}, \sigma_2^{-1}\right)$$

$$n \cdot d_{\ell_\infty}(\sigma_1, \sigma_2) \;\overset{w.h.p.}{\underset{\propto}{<}}\; d_{\ell_1}(\sigma_1, \sigma_2) \;\overset{(a)}{\underset{\propto}{<}}\; d_\tau\left(\sigma_1^{-1}, \sigma_2^{-1}\right) \;\overset{w.h.p.}{\underset{\propto}{<}}\; d_{\mathbf{x},\ell_1}\left(\sigma_1^{-1}, \sigma_2^{-1}\right)$$

- (a): [Diaconis 1977]
- $\underset{\propto}{<}$: less than, after the right hand side is scaled by some constant
- w.h.p.: when σ_1 is drawn uniformly from $\mathcal{S}_n$

In general

$$\frac{1}{n-1} d_\tau(\sigma_1, \sigma_2) \leq d_{\mathbf{x},\ell_1}(\mathbf{x}_{\sigma_1}, \mathbf{x}_{\sigma_2}) \leq d_\tau(\sigma_1, \sigma_2).$$

With high probability

For any $c < 1/2$, when σ_1 is uniformly drawn from $\mathcal{S}_n$,

$$c \cdot d_\tau(\sigma_1, \sigma_2) \leq d_{\mathbf{x},\ell_1}(\sigma_1, \sigma_2) \quad w.h.p.$$

Probabilistic argument:

$$\mathbb{E}[X_\tau] \approx \frac{n^2}{4} \qquad \mathrm{Var}[X_\tau] \approx \frac{n^3}{36}$$

$$\mathbb{E}[X_{\mathbf{x},\ell_1}] > \frac{n^2}{8} \qquad \mathrm{Var}[X_{\mathbf{x},\ell_1}] < \frac{n^3}{3}$$

For $c < 1/2$, applying Chebyshev's inequality,

$$\mathbb{P}[c \cdot X_\tau > X_{\mathbf{x},\ell_1}] = O(1/n).$$

Implication

For distortion measures d and d', if

$$d'(\sigma_1, \sigma_2) \underset{\propto}{<} d(\sigma_1, \sigma_2),$$

then under both average-case and worst-case distortion,

$$\text{a } (n, D_n) \text{ code for } \mathcal{X} \; (\mathcal{S}_n, d) \quad \Rightarrow \quad \text{a } (n, c \cdot D_n) \text{ code for } \mathcal{X} \; \big(\mathcal{S}_n, d'\big)$$

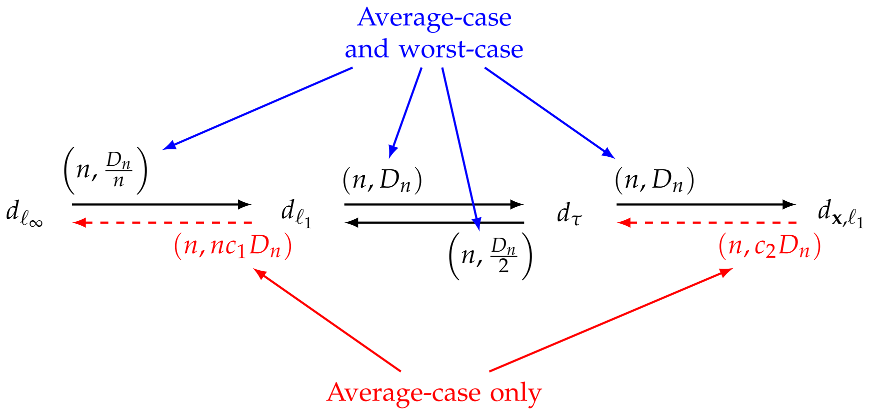
Implication

For distortion measures d and d', if

$$d'(\sigma_1, \sigma_2) \overset{w.h.p.}{\lesssim_{\propto}} d(\sigma_1, \sigma_2),$$

then under average-case distortion,

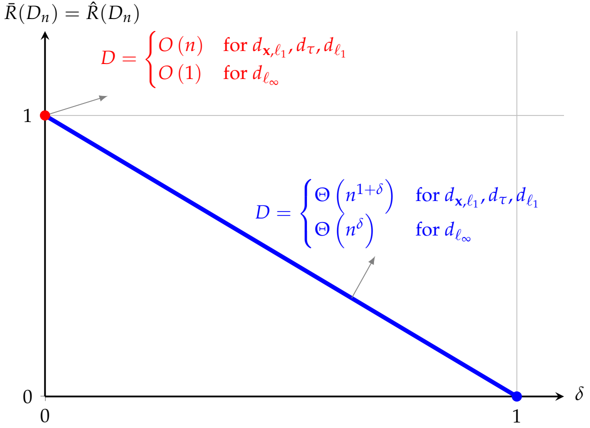
$$\text{a } (n, D_n) \text{ code for } \mathcal{X}\ (\mathcal{S}_n, d) \quad \Rightarrow \quad \text{a } (n, c \cdot D_n) \text{ code for } \mathcal{X}\ \left(\mathcal{S}_n, d'\right)$$



Average-case
and worst-case
$\left(n, \frac{D_n}{n}\right)$
(n, D_n)
(n, D_n)
d_{ℓ_∞}
d_{ℓ_1}
d_τ
$d_{\mathbf{x},\ell_1}$
$(n, nc_1 D_n)$
$\left(n, \frac{D_n}{2}\right)$
$(n, c_2 D_n)$
Average-case only

Rate distortion functions

- [**W.**, Mazumdar & Wornell, ISIT'13]: worst-case d_τ and $d_{\mathbf{x},\ell_1}$
- More analysis
  - more distortion measures
  - worst-case and average-case RDFs identical



Lossy compression schemes

- ℓ_1 and ℓ_∞ distance of permutation vectors
  - quantize by sorting subsequences that corresponding to a range of ranking
  - Time complexity: $O(n \log n)$
- Kendall tau distance
  - quantization by sorting subsequences
  - Time complexity: $O(n \log n)$

$$\underbrace{\boxed{\sigma_1, \sigma_2, \ldots, \sigma_m}}_{\text{sort}}\underbrace{\boxed{\sigma_{m+1}, \ldots, \sigma_{2m}}}_{\text{sort}}\boxed{\cdots}\underbrace{\boxed{\sigma_{(k-1)m+1}, \ldots, \sigma_{km}}}_{\text{sort}}\underbrace{\boxed{\sigma_{km+1}, \ldots, \sigma_n}}_{\text{keep}}$$

- ℓ_1 distance of inversion vectors
  - component-wise scalar quantization
  - Time complexity: $O(n)$

Discussions

- A lossy compression scheme for one distortion measure effectively preserves distortion under other measures considered in this talk
- RDF holds for any error criterion between average-case distortion and worst-case distortion
  - Example:

$$\lim_{n\to\infty} \mathbb{P}\left[d(f_n(\sigma),\sigma) > D_n\right] = 0$$

Future directions

- More distortion measures: correspond to top-k selection
- More source models: non-uniform distrition over $\mathcal{S}_n$
  - Mallows model
  - …