A Rate-Distortion Approach to Information Relevance

Saurabh Shintre Tiago Vinhoza Joao Barros Sidharth Jaggi

Novelty of Information Theory

- Information as random process
- Communication as reproduction of the process at some other space-time point
- Loss-less or lossy
- Limits on storage and transmission capacities

Loss-less storage

- Entropy
- Asymptotic limit of loss-less compression
- Other measures: Renyi entropy, ε-δ entropy
- Further compression if loss is allowed

How to define loss ?

- Distortion function (Hamming, Mean-square)
- Cost of misrepresentation
- Compress so that the overall cost is limited (or the other way around)

Rate-Distortion Theory

- Bounded distortion function

$$d(x,\hat{x}) \in \Re^{+}$$

- Distortion between sequences

$$d(X,\hat{X}) = \frac{1}{n}\sum_{j=1}^{n} d(x_j,\hat{x}_j)$$

$\hat{X}$ is the reproduction sequence

- Overall distortion of code is the average distortion overall source sequences.

$$d(\chi,\hat{\chi}) = \underset{p(\mathrm{X},\hat{\mathrm{X}})}{E} d(X,\hat{X})$$

Rate-Distortion Code

- A Rate-Distortion pair (R,D) is said to be achievable if there exists a $(2^{nR}, n)$ code (f_n, g_n) such that:

$$\lim_{n\to\infty} Ed(X, g_n(f_n(X))) \le D$$

- Rate-distortion function $R(D)$ is the infimum of all achievable rates R for a given distortion D

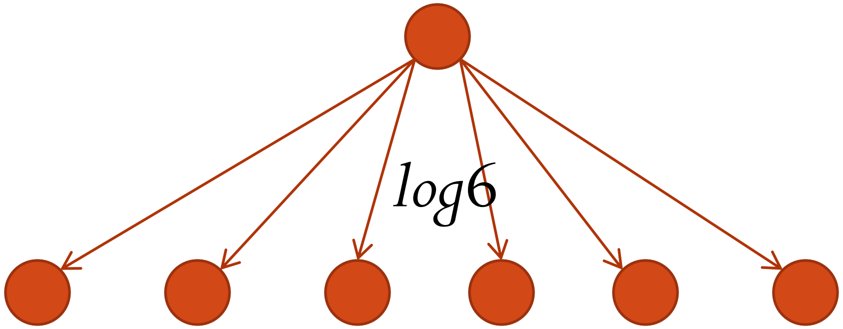
Rate-Distortion Theorem

$$R(D) = \min_{p(\hat{x}|x) : \sum_{(x,\hat{x})} p(x)\, p(\hat{x}|x)\, d(x,\hat{x}) \le D} I(X;\hat{X})$$

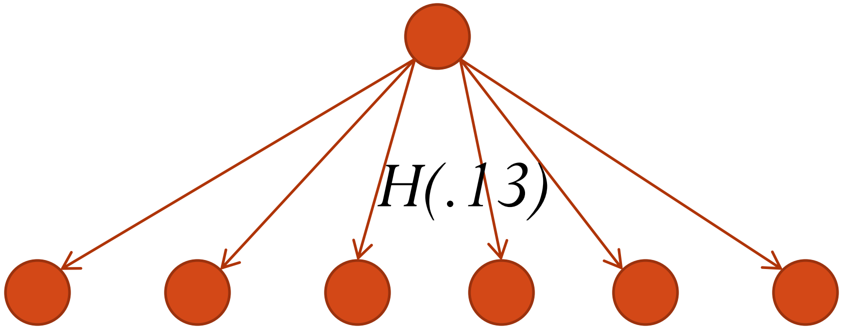
Is Relevance the same as R-D ?

- Consider a dice throwing experiment.
- Bettor i bets on the number i from $\{1,2,3,4,5,6\}$ and wins if i is the outcome.
- Only the event $X=i$ *or* $X=i^C$ is of relevance.
- Consider uniform distribution.

Comparison of methods



(a) Equal relevance



(b) Unequal relevance

Intuition

- This happens as each user *i* can incur error in the outcomes other than *i*.

- We define relevance of an outcome as the error that can be incurred in its representation.

- Different from R-D as per-letter criteria.

Mathematical modeling

- Bounded distortion function $d(x,\hat{x})$

- $I(x,\ \omega_i)$ is a binary indicator variable for each outcome ω_i

- Distortion between a sequence and its reproduction is:

$$d_i(X,\hat{X}) = \frac{1}{N_i(X)} \sum_{j=1}^{n} d(x_j,\hat{x}_j).I(x_j,\omega_i)$$

- $N_i(X)$ is the number of occurrences of ω_i *in* X

Rate-Relevance Code

- A Rate-Relevance pair $(R, e_1, e_2, \ldots, e_m)$ is said to be achievable if there exists a $(2^{nR}, n)$ code (f_n, g_n) such that :

$$\lim_{n\to\infty} Ed_i(X, g_n(f_n(X))) \leq e_i \qquad \forall i \in \{1,2,\ldots,m\}$$

- Rate-relevance function $R(e_1, e_2, \ldots, e_m)$ is the infimum of all achievable rates for a given relevance vector $\{e_1, e_2, \ldots, e_m\}$.

Claim:

$$R(e_1, e_{2,}, \ldots, e_m) = \min_{p(\hat{x}|x): \sum_{(x,\hat{x})} p(x,\hat{x}) d(x,\hat{x}) I(x,\omega_i) \le p_i e_i \quad \forall i} I(X;\hat{X})$$

We already have the converse; achievability needs further work.

Difference with R-D

- Per letter criteria gives higher control over distortion.

- Some applications may not allow error in a particular outcome, which implies the cost corresponding to that outcome should be kept infinite, but R-D theory has problems with unbounded distortion functions. (I guess ☺)

- What is the channel coding equivalent?

The way ahead

- Finish this ☺
- Performance with joint representation and meaning of terms like independent relevance.
- Application to channel coding. (UEP exists but doesn't seem right)
- Non-linear constraints

Questions for me?

- Mathematical inaccuracies
- Philosophical questions (why am I doing this?)
- Where does this apply to?

Questions for the guests

- Does it make sense?
- Has it been thought about/solved already?
- Issues, if any?
- Does it REALLY make sense?