Lempel–Ziv Compression (§6)

Last time: Symbol codes ($C: \mathcal{A} \to \{0,1\}^+$), Kraft's inequality ($\sum_x 2^{-\ell_x} \leq 1$), $H(X) \leq L(X,C) < H(X)+1$ for UD codes, achievable by $\ell_x = \lceil \log \frac{1}{P(x)} \rceil$

optimal code via Huffman algo

$\boxed{\text{Today:}}$ Compression algos that operate on "stream" of symbols, can emit <1 bit/symbol, are asymptotically optimal for IID sources ($R \to H(X)$), but are also adaptive!

Used in GIF, ogg
Similar to LZ78 which is used in ZIP, PNG, ...
(together with Huffman)

Lempel–Ziv's Coding algo (LZ78):

assume: stream ends with special symbol $\perp$

$*$ phrases $\leftarrow [\emptyset]$ $\to$ $\boxed{\text{HW}}$

$*$ while more to compress:

- read symbols until we obtain "phrase" $\pi \notin$ phrases

  $\Rightarrow \pi = \boxed{\tau}\boxed{x}$ where $\tau \in$ phrases, $x \in \mathcal{A}$
- append π to phrases
- $k \leftarrow$ index of τ in phrases
- write (k, x) in bits

Splits input into minimal distinct "phrases"

use $\lceil \log(j) \rceil$ bits in j-th step ($j = 1, 2, \ldots$)

Can skip if $x = \perp$ (last step)

Example: Let's compress A|B|BA|BAA|BAAB|AB|A$\perp$:

Step	0	1	2	3	4	5	6	7
phrases	$\emptyset$	A	B	BA	BAA	BAAB	AB	A$\perp$
(k,x)	—	(0,A)	(0,B)	(2,A)	(3,A)	(4,B)	(1,B)	(1,$\perp$)
Compression		–,0	0,1	10,0	11,0	100,1	001,1	001,–
#bits for k		0	1	2	2	3	3	3

$\Rightarrow$ ☹ 14 bits compressed into 20 bits... but the principle is sound ☺

Q: Intuition? Clear how to decompress?

Analysis? Let $R = \frac{\ell}{n}$ the compression rate, where $\ell = \#$bits of compression.

* Average case: For IID source, where $X^N = X_1 \cdots X_N \overset{IID}{\sim} P$:

$$E[R] \leq H(P) + O\left(\frac{1}{\log N}\right) \longrightarrow H(P)$$

* Worst case: For any string $x^N = x_1 \cdots x_N$,

$$R \leq \log \#\mathcal{A} + O\left(\frac{1}{\log N}\right) \longrightarrow \log \#\mathcal{A}$$

$f = O(g)$ means $\exists C > 0 : f(N) \leq C \cdot g(N) \ \forall N$

Warmup: Fix x^N and assume LZ decomposes it into c phrases:

$$x^N = x_1 \cdots x_N = \pi_1 \cdots \pi_c$$

disjoint phrases

$$\Rightarrow \ell = \sum_{j=1}^{c} \left(\lceil \log(j) \rceil + \lceil \log \#\mathcal{A} \rceil \right)$$
$$\leq c \cdot \log(c) + c\left(1 + \lceil \log \#\mathcal{A} \rceil\right)$$

need to understand

To make progress, we need to upper-bound c. For worst-case analysis, we would try to bound c in terms of N $\longrightarrow$ $\boxed{\text{EX CLASS}}$.
We focus on the average case, so want to relate c to $P(x^N)$!

For Simplicity: $\boxed{\text{Assume all } P(x) \leq \frac{1}{2}}$ ← but arbitrary $\#\mathcal{A}$:)

For our fixed string x^N, consider:

$$\Pi_k = \{ \pi_i \mid 2^{-k-1} < P(\pi_i) \leq 2^{-k} \}$$ ← classify phrases according to probability

* for any phrase: $P(\pi) = \Pr(Y^N \text{ has prefix } \pi)$

* any Y^n has at most one prefix in Π_k (if both π_i & π_j are prefix then $\pi_i = \pi_j * \cdots *$ (or vice versa) $\Rightarrow P(\pi_i) \leq P(\pi_j) \frac{1}{2}$ ↯)

$$\Rightarrow 1 \geq \Pr(Y^N \text{ has prefix in } \Pi_k) \geq \sum_{\pi \in \Pi_k} P(\pi) \geq \#\Pi_k \cdot 2^{-k-1}$$

Thus: $\#\Pi_k \leq 2^{k+1}$

How large can $P(x^N)$ be if we know it has c phrases?

$$P(x^N) = \prod_k \prod_{\pi \in \Pi_k} P(\pi)$$ maximal if 2^{k+1} phrases in Π_k $(\forall k)$

$$\leq (2^{-0})^{2^{0+1}} (2^{-1})^{2^{1+1}} \cdots (2^{-(L-1)})^{2^{L}} (2^{-L})^{c-(2^{L+1}-2)}$$

where L is maximal with $\sum_{k=0}^{L} 2^k = 2^{L+1} - 2 \leq c$. Note:

① $c \geq 2^{L+1} - 2 \implies L \leq \log(c+2) - 1 \leq \log(c)$ if $c \geq 2$, i.e. $N > 1$

② $c < 2^{L+2} - 2 < 2^{L+2} \implies L \geq \log(c) - 2$

Thus:

$$\log \frac{1}{P(x^N)} \geq \underbrace{\sum_{k=1}^{L} (k-1) 2^k}_{\text{check by induction } =\ (L-2)2^{L+1} + 4} + L(c - 2^{L+1} + 2)$$

$$= (L-2)2^{L+1} + 4 + L(c - 2^{L+1} + 2)$$

$$= -4 2^L + 4 + cL + 2L$$

$$\overset{①}{\underset{②}{\geq}} -4c + \cancel{4} + c(\log c - 2) + 2(\log c - \cancel{2})$$

$$= c \cdot \log c - 6c$$

Take expectation values and use $H(X^N) = N \cdot H(P)$:

$$\boxed{N \cdot H(P) \geq E[c \cdot \log c] - 6 E[c]} \quad (*)$$

$$\implies E[R] = \frac{1}{N} E[\ell] \overset{\text{warmup}}{\leq} \frac{1}{N} E[c \cdot \log c] + \frac{1}{N}(\lceil \log \#\mathcal{A} \rceil + 1) \cdot E[c]$$

$$\leq H(P) + O\Big(\underbrace{\tfrac{1}{N} E[c]}_{\text{want that} \to 0}\Big)$$

How to deal with $E[c]$?

$E[C]\log E[C] \leq E[C\cdot\log C] \overset{(*)}{\leq} (H(P)+6)N$ since certainly $C \leq N$

Jensen: $f(x) = x\cdot\log x$ is convex

... so $E[C]$ has to grow slower than linear? In fact:

$\Longrightarrow E[C] = O\left(\frac{N}{\log N}\right)$ and so we arrive at

$\Longrightarrow E[R] \leq H(P) + O\left(\frac{1}{\log N}\right)$

Nice!

Why is this true? Assume that $f(N)\cdot\log f(N) \leq \gamma\cdot N$ for large N.

We claim that $f(N) < (\gamma+1)\frac{N}{\log N}$. Indeed, otherwise we have $f(N) \geq (\gamma+1)\frac{N}{\log N}$ for a subsequence of $N\to\infty$. Then:

$$f(N)\cdot\log f(N) \geq (\gamma+1)\frac{N}{\log N}\log\Big(\underbrace{(\gamma+1)}_{\geq 1}\frac{N}{\log N}\Big)$$

$$\geq (\gamma+1)N\Big(1 - \underbrace{\frac{\log\log N}{\log N}}_{\to 0}\Big)$$

$\lightning$