(F1X) Teoria dell'Informazione e della Trasmissione

Primo Teorema di Shannon

Docente: *Nicolò Cesa-Bianchi* versione 19 giugno 2017

La modellizzazione statistica della sorgente permette di interpretare le parole un codice sorgente come descrizioni compresse delle realizzazioni di una variabile casuale X associata alla sorgente. Il prossimo risultato rivela che i codici di Shannon forniscono una descrizione delle realizzazioni di X di lunghezza media quasi ottimale rispetto a tutti i codici istantanei.

Fatto 1 *Per ogni sorgente $\langle \mathcal{X}, p \rangle$ con $\mathcal{X} = \{x_1, \ldots, x_m\}$ e $p = (p_1, \ldots, p_m)$. Dato il codice istantaneo di Shannon c con lunghezze $\ell_i = \ell_c(x_i)$ tali che $\ell_i = \left\lceil \log_D \frac{1}{p_i} \right\rceil$ per $i = 1, \ldots, m$, vale*

$$\mathbb{E}\big[\ell_c\big] < H_D(X) + 1 \ .$$

Dimostrazione.

$$\mathbb{E}\big[\ell_c\big] = \sum_{i=1}^{m} p_i \left\lceil \log_D \frac{1}{p_i} \right\rceil < \sum_{i=1}^{m} p_i \left(\log_D \frac{1}{p_i} + 1 \right) = H_D(X) + 1 \ .$$

□

Quindi, combinando questo risultato con $\mathbb{E}\big[\ell_c\big] \geq H_D(X)$ che vale per ogni codice istantaneo c, otteniamo che l'entropia è il numero minimo medio di bit necessario e sufficiente (a meno di una costante additiva non maggiore di 1) per descrivere il valore di una variabile casuale con un codice binario istantaneo.

Come abbiamo appena visto, i codici di Shannon sono codici istantanei quasi ottimi, nel senso che in media una parola di codice è lunga al più un simbolo in più del necessario. D'altra parte, l'inefficienza dei codici di Shannon cresce linearmente con la lunghezza del messaggio da codificare. Infatti, la lunghezza della codifica di Shannon di un messaggio $(x_1, \ldots, x_n)$, generato con n estrazioni da una sorgente $\langle \mathcal{X}, p \rangle$, è pari a

$$\sum_{i=1}^{n} \left\lceil \log_D \frac{1}{p(x_i)} \right\rceil \ .$$

Possiamo incrementare l'efficienza di questi codici usando una tecnica nota come **codifica a blocchi**. Dimostriamo che se codifichiamo con Shannon a blocchi, allora la lunghezza media della parola di codice per simbolo sorgente è asintotica all'entropia quando la lunghezza del blocco cresce all'infinito. Questo è il risultato noto come primo teorema di Shannon, o *source coding theorem.*

La codifica a blocchi suddivide ogni messaggio in blocchi di simboli sorgente dove ogni blocco ha la stessa lunghezza. I blocchi vengono quindi codificati con un codice per la sorgente i cui simboli sono tutti i possibili blocchi di lunghezza data. Per prima cosa creiamo un modello di sorgente a blocchi a partire dal modello $\langle \mathcal{X}, p \rangle$. Ogni blocco di n simboli sorgente sia generato mediante n estrazioni indipendenti e identicamente distribuite (i.i.d.) secondo p. Questo definisce il modello

$\langle \mathcal{X}^n, P_n \rangle$, dove $\mathcal{X}^n$ è l'insieme delle n-uple $(x_1, \ldots, x_n)$ di simboli di $\mathcal{X}$ e P_n è la distribuzione su $\mathcal{X}^n$ associata a n estrazioni i.i.d. secondo p. Per l'ipotesi di indipendenza,

$$P_n(x_1, \ldots, x_n) = \prod_{i=1}^{n} p(x_i) \ .$$

Vediamo ora come esprimere l'entropia di P_n in termini dell'entropia di p. Sia X una variabile casuale con distribuzione p e $X_1, \ldots, X_n$ variabili casuali i.i.d. anch'esse con distribuzione p. Abbiamo

$$\begin{aligned} H(X_1, \ldots, X_n) &= \sum_{x_1, \ldots, x_n} P_n(x_1, \ldots, x_n) \log_2 \frac{1}{P_n(x_1, \ldots, x_n)} \\ &= \sum_{x_1} \cdots \sum_{x_n} \left(\prod_{i=1}^{n} p(x_i) \right) \sum_{i=1}^{n} \log_2 \frac{1}{p(x_i)} \\ &= \sum_{i=1}^{n} \sum_{x_i} p(x_i) \log_2 \frac{1}{p(x_i)} \\ &= n\, H(X) \ . \end{aligned}$$

Cioè l'entropia della sorgente a blocchi di n simboli è n volte l'entropia della sorgente base.

Siamo ora pronti per enunciare e dimostrare il teorema.

Teorema 2 (Primo teorema di Shannon) *Sia $C_n : \mathcal{X}^n \to \mathcal{D}^+$ un codice di Shannon D-ario a blocchi per la sorgente $\langle \mathcal{X}, p \rangle$. Ovvero,*

$$\ell_{C_n}(x_1, \ldots, x_n) = \left\lceil \log_D \frac{1}{P_n(x_1, \ldots, x_n)} \right\rceil \ .$$

Allora

$$\lim_{n \to \infty} \frac{1}{n} \mathbb{E}\big[\ell_{C_n}\big] = H_D(X)$$

dove $H_D(X)$ indica l'entropia della sorgente $\langle \mathcal{X}, p \rangle$.

Dimostrazione. Per prima cosa osserviamo che siccome C_n è un codice istantaneo su una sorgente $\langle \mathcal{X}^n, P_n \rangle$ vale il minorante $\mathbb{E}\big[\ell_{C_n}\big] \geq H_D(X_1, \ldots, X_n)$ ed il maggiorante $\mathbb{E}\big[\ell_{C_n}\big] < H_D(X_1, \ldots, X_n) + 1$. Quindi possiamo scrivere

$$n\, H_D(X) = H_D(X_1, \ldots, X_n) \leq \mathbb{E}\big[\ell_{C_n}\big] < H_D(X_1, \ldots, X_n) + 1 = n\, H_D(X) + 1 \ .$$

Dividendo entrambi i membri per n otteniamo

$$H_D(X) \leq \frac{1}{n} \mathbb{E}\big[\ell_{C_n}\big] < H_D(X) + \frac{1}{n} \ .$$

Da quest'ultima relazione ricaviamo immediatamente l'enunciato del teorema. □

Ora che abbiamo associato l'entropia alla lunghezza minima media di codici istantanei, diamo un analogo significato operativo all'entropia relativa, mostrando che essa corrisponde alla differenza fra la lunghezza media di un codice di Shannon costruito sul modello di sorgente e quello di un codice di Shannon costruito su un diverso modello di sorgente.

Teorema 3 *Dato un modello di sorgente $\langle \mathcal{X}, p \rangle$, se $c : \mathcal{X} \to \mathcal{D}^+$ è un codice di Shannon con lunghezze $\ell_c(x) = \lceil \frac{1}{q(x)} \rceil$, dove q è una distribuzione arbitraria su $\mathcal{X}$, allora*

$$\mathbb{E}[\ell_c] < H_D(X) + D(X\|Y) + 1$$

dove X ha distribuzione p e Y ha distribuzione q.

Dimostrazione. È sufficiente osservare che

$$\begin{aligned}
\mathbb{E}[\ell_c] &= \sum_{x \in \mathcal{X}} p(x) \left\lceil \log_D \frac{1}{q(x)} \right\rceil \\
&< \sum_{x \in \mathcal{X}} p(x) \log_D \frac{1}{q(x)} + 1 \\
&= \sum_{x \in \mathcal{X}} p(x) \log_D \left(\frac{1}{q(x)} \frac{p(x)}{p(x)} \right) + 1 \\
&= \sum_{x \in \mathcal{X}} p(x) \log_D \frac{p(x)}{q(x)} + \sum_{x \in \mathcal{X}} p(x) \log_D \frac{1}{p(x)} + 1 \\
&= D(X\|Y) + H_D(X) + 1 \ .
\end{aligned}$$

□