(F1X) Teoria dell'Informazione e della Trasmissione

9 – Codifica di canale

Docente: *Nicolò Cesa-Bianchi* versione 6 aprile 2016

Ricordiamo l'obiettivo generale della codifica canale: vogliamo trasmettere un messaggio attraverso un canale con rumore massimizzando la quantità di informazione trasmessa per uso del canale e simultaneamente minimizzando la probabilità di errore di decodifica. Il primo ed il secondo teorema di Shannon mostrano come sia possibile trasmettere la massima quantità di informazione (che è finita, in quanto limitata dall'entropia della sorgente) con probabilità di errore arbitrariamente vicina a zero. Operativamente, Shannon dimostra che è sufficiente usare una codifica a due stadi: prima codifico il messaggio con un codice sorgente ottimo. Quindi ricodifico la parola di codice ottenuta tramite un codice canale ottimo.

Formalmente, un **canale discreto senza memoria** è definito da $\langle \mathcal{X}, \mathcal{Y}, p(y \mid x) \rangle$ dove $\mathcal{X}$ è un alfabeto finito di simboli di ingresso, $\mathcal{Y}$ è un insieme finito di simboli di uscita e $p(y \mid x)$ è la matrice di canale, ovvero la matrice delle probabilità condizionate $p(y \mid x) = \mathbb{P}(Y = y \mid X = x)$ di ricevere il simbolo $y \in \mathcal{Y}$ dato che è stato inviato $x \in \mathcal{X}$. I simboli di ingresso e uscita possono anche appartenere a insiemi diversi, ovvero $\mathcal{X} \neq \mathcal{Y}$. Per esempio, un simbolo di ingresso inviato potrebbe essere sostituito con un altro dal canale.

Qui sopra sono indicati due semplici esempi di canale. Sono entrambi canali binari, $\mathcal{X} = \mathcal{Y} = \{0, 1\}$. Le probabilità sulle frecce indicano la matrice $p(y \mid x)$. Il primo è un canale binario senza rumore: $p(1 \mid 1) = p(0 \mid 0) = 1$ e $p(1 \mid 0) = p(0 \mid 1) = 0$. Il secondo è un canale binario con rumore simmetrico di parametro p, $p(1 \mid 1) = p(0 \mid 0) = 1 - p$ e $p(1 \mid 0) = p(0 \mid 1) = p$.

Un canale è detto senza memoria per la seguente ragione. Supponiamo che il canale venga usato n volte per inviare il messaggio $x^n = (x_1, \ldots, x_n)$ ottenendo in uscita $y^n = (y_1, \ldots, y_n)$. Per la chain rule delle probabilità,

$$p(y^n \mid x^n) = p(y_n \mid y^{n-1}, x^n) p(y_{n-1} \mid y^{n-2}, x^n) \times \cdots \times p(y_1 \mid x^n) \ .$$

Ora, se il canale è senza memoria vale che

$$\begin{aligned} p(y_n \mid y^{n-1}, x^n) &= p(y_n \mid x_n) \\ p(y_{n-1} \mid y^{n-2}, x^n) &= p(y_{n-1} \mid x_{n-1}) \\ &\vdots \\ p(y_1 \mid x^n) &= p(y_1 \mid x_1) \ . \end{aligned}$$

Ovvero,

$$p(y^n \mid x^n) = \prod_{t=1}^{n} p(y_t \mid x_t) \ .$$

In altri termini, un canale è senza memoria quando l'uscita ottenuta ad ogni uso del canale, condizionata sull'ingresso, è indipendente dagli utilizzi passati e futuri.

Diamo ora la definizione di capacità di una canale. Il secondo teorema di Shannon fornirà un'interpretazione operativa della capacità come massima quantità di informazione trasmissibile ad ogni uso del canale quando la probabilità di errore di decodifica è arbitrariamente piccola.

La **capacità** di un canale $\langle \mathcal{X}, \mathcal{Y}, p(y \mid x) \rangle$ è definita come

$$C = \max_{p(x)} I(X, Y)$$

dove il massimo è su tutte le distribuzioni di probabilità sui simboli di ingresso.

Enunciamo alcune semplici proprietà della capacità. Ricordando che

$$0 \leq I(X, Y) = \begin{cases} H(X) - H(X \mid Y) \\ H(Y) - H(Y \mid X) \end{cases}$$

e che $H(X) \leq \log_2 |\mathcal{X}|$, $H(Y) \leq \log_2 |\mathcal{Y}|$, otteniamo che

$$0 \leq C \leq \min\{\log_2 |\mathcal{X}|, \log_2 |\mathcal{Y}|\} \ .$$

Calcoliamo la capacità di alcuni canali discreti senza memoria cominciando dal caso più semplice: il canale binario senza rumore

$$0 \xrightarrow{1} 0$$

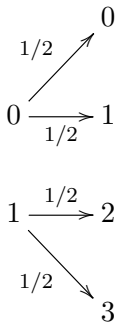
$$1 \xrightarrow{1} 1$$

Dato che $H(X \mid Y) = 0$, cioè dato Y in uscita non ho incertezza su X in entrata, dalla definizione di capacità otteniamo

$$C = \max_{p(x)} I(X, Y) = \max_{p(x)} \Big(H(X) - H(X \mid Y) \Big) = \max_{p(x)} H(X) = 1 \ .$$

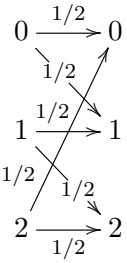
Si noti che 1 bit è anche la massima quantità di informazione trasmissibile ad ogni uso del canale, ovvero in questo caso semplice l'interpretazione operativa della capacità è evidente.

Passiamo ad un secondo esempio semplice, il canale con rumore e uscite disgiunte. Qui l'alfabeto d'ingresso è $\{0, 1\}$ mentre quello d'uscita è $\{0, 1, 2, 3\}$. La matrice di canale è data da



Come per il canale binario senza rumore, $H(X \mid Y) = 0$ e quindi $C = 1$. Anche in questo caso 1 bit è evidentemente la massima quantità di informazione trasmissibile ad ogni uso del canale.

Un canale lievemente più complesso è la "macchina da scrivere rumorosa", dove l'alfabeto di ingresso e quello di uscita sono entrambi $\{0, 1, 2\}$ e la matrice di canale è definita da

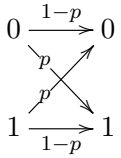


In questo caso è immediato vedere che per qualunque simbolo di ingresso x, vale $H(Y \mid X = x) = 1$. Quindi

$$C = \max_{p(x)} I(X, Y) = \max_{p(x)} \Big(H(Y) - H(Y \mid X) \Big) = \max_{p(x)} H(Y) - 1 = \log_2 3 - 1$$

dove abbiamo utilizzato il fatto (facile da verificare) che, in questo canale, ad una distribuzione $p(x)$ uniforme in ingresso corrisponde una distribuzione $p(y)$ uniforme in uscita.

Terminiamo con due esempi più complessi. Per primo, il canale binario simmetrico



Iniziamo coll'osservare

$$I(X, Y) = H(Y) - H(Y \mid X) = H(Y) - H(Y \mid X = 0)\mathbb{P}(X = 0) - H(Y \mid X = 1)\mathbb{P}(X = 1) \ .$$

Vediamo allora che

$$\begin{aligned} H(Y \mid X = 0) &= -\mathbb{P}(Y = 0 \mid X = 0) \log_2 \mathbb{P}(Y = 0 \mid X = 0) - \mathbb{P}(Y = 1 \mid X = 0) \log_2 \mathbb{P}(Y = 1 \mid X = 0) \\ &= -(1 - p) \log_2 (1 - p) - p \log_2 p \\ &= H(p) \end{aligned}$$

dove $H(p)$, lo ricordiamo, è l'entropia di una Bernoulliana di parametro p. In modo del tutto analogo si dimostra che $H(Y \mid X = 1) = H(p)$. Quindi,

$$C = \max_{p(x)} I(X, Y) = \max_{p(x)} H(Y) - H(p) \ .$$

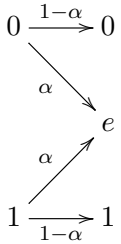
Ci resta quindi da calcolare il massimo di $H(Y)$ al variare di $p(x)$. Se scriviamo

$$\begin{aligned} \mathbb{P}(Y = 1) &= \mathbb{P}(Y = 1 \mid X = 0)\mathbb{P}(X = 0) + \mathbb{P}(Y = 1 \mid X = 1)\mathbb{P}(X = 1) \\ &= p\, \mathbb{P}(X = 0) + (1 - p)\, \mathbb{P}(X = 1) \end{aligned}$$

notiamo che quando $\mathbb{P}(X=1)=\frac{1}{2}$ abbiamo che $\mathbb{P}(Y=1)=\frac{1}{2}$. Quindi, per questa scelta di $p(x)$ abbiamo che $H(Y)=1$, che è massima dato che $\mathcal{Y}=\{0,1\}$. Possiamo quindi concludere che

$$C = 1 - H(p) \ .$$

L'ultimo esempio di canale è quello binario a cancellazione, con $\mathcal{X}=\{0,1\}$, $\mathcal{Y}=\{0,1,e\}$ e matrice di canale definita come



Il simbolo "e" nell'alfabeto di uscita rappresenta la cancellazione di un simbolo di ingresso.

Cominciamo coll'osservare che $H(Y \mid X=0) = H(Y \mid X=1) = H(\alpha)$ e perciò $H(Y \mid X) = H(\alpha)$. Questo implica che $I(X,Y) = H(Y) - H(\alpha)$ e, come nell'esempio precedente, dobbiamo calcolare il massimo di $H(Y)$ al variare di $p(x)$.

Introduciamo la variabile casuale Bernoulliana

$$Z = \begin{cases} 1 & \text{se } Y = e \\ 0 & \text{altrimenti.} \end{cases}$$

e notiamo che

$$H(Y,Z) = H(Y) + \underbrace{H(Z \mid Y)}_{=0} = H(Y)$$
$$H(Y,Z) = H(Z) + H(Y \mid Z)$$

quindi $H(Y) = H(Z) + H(Y \mid Z)$. Per calcolare $H(Z)$ basta osservare che

$$\begin{aligned} \mathbb{P}(Z=1) &= \mathbb{P}(Z=1 \mid X=0)\mathbb{P}(X=0) + \mathbb{P}(Z=1 \mid X=1)\mathbb{P}(X=1) \\ &= \alpha\,\mathbb{P}(X=0) + \alpha\,\mathbb{P}(X=1) = \alpha \ . \end{aligned}$$

Allora, evidentemente, $H(Z) = H(\alpha)$.

Per calcolare $H(Y \mid Z)$ osserviamo che $\mathbb{P}(Y=1 \mid Z=0) = \mathbb{P}(X=1)$. Quindi $H(Y \mid Z=0) = H(X)$. Possiamo allora scrivere

$$H(Y \mid Z) = H(Y \mid Z=0)\mathbb{P}(Z=0) + \underbrace{H(Y \mid Z=1)}_{=0}\mathbb{P}(Z=1) = H(X)(1-\alpha) \ .$$

Possiamo allora concludere come segue

$$\begin{aligned} C &= \max_{p(x)} H(Y) - H(\alpha) \\ &= \max_{p(x)} \Big(H(\alpha) + H(X)(1-\alpha) \Big) - H(\alpha) \\ &= (1-\alpha) \max_{p(x)} H(X) \\ &= 1 - \alpha \ . \end{aligned}$$