(F1X) Teoria dell'Informazione e della Trasmissione

13 – Teorema di Codifica Congiunta Sorgente-Canale

Docente: *Nicolò Cesa-Bianchi* versione 14 aprile 2014

In questa lezione combiniamo il primo e secondo teorema di Shannon per dimostrare il teorema di codifica congiunta sorgente-canale. Questo teorema dimostra che la codifica a due stadi, dove prima comprimiamo la sorgente e poi creiamo un codice canale per trasmettere i simboli compressi, è ottima. Questo non è scontato, infatti il codice sorgente ignora il canale, mentre il codice canale ignora la sorgente. Invece, il teorema dimostra che non c'è modo di avere una codifica più efficiente anche mettendo insieme sorgente e canale.

I due teoremi di Shannon mostrano che:

1. possiamo comprimere senza perdita di informazione ad un tasso arbitrariamente vicino all'entropia per simbolo sorgente,
2. possiamo trasmettere con probabilità di errore arbitrariamente piccola ad un tasso arbitrariamente vicino alla capacità per uso del canale.

Entropia e capacità forniscono condizioni necessarie e sufficienti: non possiamo comprimere ad un tasso superiore all'entropia e trasmettere ad un tasso superiore alla capacità. Ora dimostriamo che possiamo trasmettere un simbolo sorgente per uso del canale con errore medio tendente a zero se e solo se l'entropia della sorgente è minore della capacità del canale.

Intuitivamente, questo avviene grazie alla proprietà AEP: se consideriamo una sorgente $\langle \mathcal{V}, p(v) \rangle$ di entropia H, sappiamo che il numero di sequenze di n simboli sorgente sulle quali si concentra la probabilità è circa 2^{nH}. Questo numero coincide col numero di messaggi che vogliamo vengano decodificati correttamente con alta probabilità, in quanto —per AEP— i rimanenti messaggi hanno una probabilità molto bassa di essere estratti dalla sorgente. Se usiamo il canale n volte per trasmettere uno di questi 2^{nH} messaggi, abbiamo un tasso di trasmissione pari ad H. Il secondo teorema di Shannon, garantisce allora una probabilità di errore di decodifica asintoticamente nulla quando $H < C$.

Teorema 1 *Data una sorgente $\langle \mathcal{V}, p(v) \rangle$ di entropia H e dato un canale $\langle \mathcal{X}, \mathcal{Y}, p(y \mid x) \rangle$ di capacità C, si considerino codici sorgente-canale della forma*

$$v^n \to x^n(v^n) \to y^n \to \widehat{v}^n(y^n)$$

dove $x^n : \mathcal{V}^n \to \mathcal{X}^n$ è la funzione di codifica sorgente-canale e $\widehat{v}^n : \mathcal{Y}^n \to \mathcal{V}^n$ è la funzione di decodifica. Sia $p_e^{(n)} = \mathbb{P}\big(\widehat{V}^n \neq V^n\big)$ la probabilità di errore di decodifica per un blocco casuale di n simboli $V^n \in \mathcal{V}^n$ dove

$$\mathbb{P}(V^n = v^n) = p(v^n) = \prod_{i=1}^{n} p(v_i) \ .$$

Allora,

1. *Se $H < C$ allora esiste una sequenza di codici sorgente-canale tale che $p_e^{(n)} \to 0$ per $n \to \infty$. Tali codici sono costruiti concatenando un codice sorgente (indipendente dal canale) con un codice canale (indipendente dalla sorgente).*

2. *Se per una data sequenza di codici $p_e^{(n)} \to 0$ quando $n \to \infty$ allora $H \leq C$.*

DIMOSTRAZIONE. Assumiamo $H < C$ e costruiamo un codice sorgente-canale con le proprietà desiderate. Per ogni $\varepsilon > 0$ e n fissati consideriamo l'insieme $A_\varepsilon^{(n)}$ di sequenze tipiche in $\mathcal{V}^n$. Costruiamo il codice sorgente $C_\varepsilon^{(n)} : A_\varepsilon^{(n)} \to \{0,1\}^*$ assegnando ad ogni $v^n \in A_\varepsilon^{(n)}$ una distinta parola di codice di lunghezza $\lceil \log_2 |A_\varepsilon^{(n)}| \rceil$. Mappiamo poi le restanti sequenze $v^n \notin A_\varepsilon^{(n)}$ in una stessa parola di codice arbitraria diversa dalle precedenti. Dato che la probabilità di $A_\varepsilon^{(n)}$ tende a uno per $n \to \infty$, il codice codifica soltanto le sequenze che hanno probabilità asintoticamente non zero di essere estratte dalla sorgente.

Ora applichiamo una codifica canale asintoticamente ottima all'insieme

$$M_\varepsilon^{(n)} = \left\{ C_\varepsilon^{(n)}(v^n) \,:\, v^n \in \mathcal{V}^n \right\}$$

dei messaggi (parole del codice sorgente). Riassumendo, una sequenza V^n di simboli sorgente è codificata dal codice sorgente con il messaggio binario $C_\varepsilon^{(n)}(V^n)$, quindi ricodificata dal codice canale con la sequenza X^n che in uscita dal canale diventa la sequenza Y^n. La funzione di decodifica riceve Y^n in uscita dal canale, applica la funzione di decodifica del codice canale ricavando un messaggio corrispondente ad una parola del codice sorgente, infine decodifica il messaggio utilizzando il codice sorgente e ricavando la sequenza decodificata $\widehat{V}^n$ di simboli sorgente. La probabilità di errore di decodifica è controllata nel modo seguente

$$\begin{aligned} p_e^{(n)} &= \mathbb{P}\big(\widehat{V}^n \neq V^n\big) \\ &= \mathbb{P}\big(\widehat{V}^n \neq V^n \mid V^n \in A_\varepsilon^{(n)}\big)\, \mathbb{P}\big(A_\varepsilon^{(n)}\big) + \mathbb{P}\big(\widehat{V}^n \neq V^n \mid V^n \notin A_\varepsilon^{(n)}\big)\, \mathbb{P}\big(\mathcal{V}^n \setminus A_\varepsilon^{(n)}\big) \\ &\leq \mathbb{P}\big(\widehat{V}^n \neq V^n \mid V^n \in A_\varepsilon^{(n)}\big) + \mathbb{P}\big(\mathcal{V}^n \setminus A_\varepsilon^{(n)}\big) \ . \end{aligned}$$

Dato che abbiamo usato un codice canale asintoticamente ottimo, per ogni $\varepsilon > 0$ fissato esiste n_ε tale che per ogni $n \geq n_\varepsilon$, si ha che

$$\frac{\log_2 \big|M_\varepsilon^{(n)}\big|}{n} < C \qquad \text{implica} \qquad \mathbb{P}\big(\widehat{V}^n \neq V^n \mid V^n \in A_\varepsilon^{(n)}\big) \leq \varepsilon \ .$$

Dato che $\big|M_\varepsilon^{(n)}\big| \leq 2^{n(H+\varepsilon)} + 1$ per avere

$$\frac{\log_2 \big|M_\varepsilon^{(n)}\big|}{n} < C$$

è sufficiente che $H + \varepsilon + \frac{1}{n} < C$. Inoltre, di nuovo per n abbastanza grande rispetto a ε, abbiamo che $\mathbb{P}\big(\mathcal{V}^n \setminus A_\varepsilon^{(n)}\big) \leq \varepsilon$. Quindi concludiamo che $H + \varepsilon < C$ implica $p_e^{(n)} \leq 2\varepsilon$. Infine, per l'arbitrarietà nella scelta di ε, otteniamo che $H < C$ implica $\lim_{n\to\infty} p_e^{(n)} = 0$.

Concludiamo la dimostrazione provando che se esiste una sequenza di codici sorgente-canale tale che $\lim_{n\to\infty} p_e^{(n)} = 0$ allora dev'essere $H < C$. Per prima cosa notiamo che $H = \frac{1}{n}H(V^n)$ dato che $p(v^n) = p(v_1) \times \cdots \times p(v_n)$. Poi, dato che $\widehat{V}^n$ è indipendente da V^n dato Y^n, possiamo applicare la data processing inequality alla terna $V^n, Y^n, \widehat{V}^n$. Infine, dato che Y^n è indipendente da V^n dato X^n, possiamo applicare ancora la data processing inequality alla terna V^n, X^n, Y^n. Possiamo allora scrivere

$$
\begin{aligned}
H &= \frac{1}{n}H(V^n) \\
&= \frac{1}{n}H\big(V^n \mid \widehat{V}^n\big) + \frac{1}{n}I\big(V^n, \widehat{V}^n\big) \\
&\leq \frac{1}{n}\big(1 + p_e^{(n)} n \log_2 |\mathcal{V}|\big) + \frac{1}{n}I\big(V^n, \widehat{V}^n\big) \quad \text{(per la disuguaglianza di Fano)} \\
&\leq \frac{1}{n}\big(1 + p_e^{(n)} n \log_2 |\mathcal{V}|\big) + \frac{1}{n}I\big(V^n, Y^n\big) \quad \text{(per la data processing inequality)} \\
&\leq \frac{1}{n}\big(1 + p_e^{(n)} n \log_2 |\mathcal{V}|\big) + \frac{1}{n}I\big(X^n, Y^n\big) \quad \text{(per la data processing inequality)} \\
&\leq \frac{1}{n} + p_e^{(n)} \log_2 |\mathcal{V}| + C \quad \text{(dato che la capacità massimizza l'informazione mutua).}
\end{aligned}
$$

Perciò, se $p_e^{(n)} \to 0$ dev'essere $H \leq C$, il che conclude la dimostrazione.

□