Appunti di TEORIA DELL'INFORMAZIONE

Sistemi di elaborazione e trasmissione delle informazioni
Prof. Francesco Taurisano

Indice

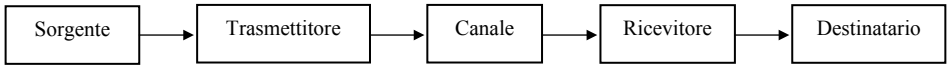
Introduzione

La teoria dell'informazione studia le tecniche di trasferimento nel tempo (immagazzinamento) e nello spazio (trasmissione) di messaggi a cui è legato il concetto di informazione. Tale teoria risale e si ricollega al concetto aristotelico di *"dar forma"*; l'etimologia del termine *informazione* è riconducibile al termine latino *"forma"* ed è connessa a quella del termine formaggio *"informaticum"*, cioè messo in forma.
Affinché si possa parlare di informazione occorre che esistano delle differenze. Se tutto fosse uniforme non esisterebbe informazione e quindi non sarebbe possibile comunicare.
La formalizzazione rigorosa della teoria dell'informazione è dovuta a Claude Elwood Shannon, un ingegnere della compagnia telefonica Bell, poi trasferitosi al centro di ricerca del Massachusetts Institute of Technology, che nel 1948 pubblicò un lavoro di ricerca innovativo, rilevatosi fondamentale per tutti gli sviluppi successivi.
L'obiettivo della teoria dell'informazione è capire come deve essere rappresentato un messaggio prodotto da una sorgente (codifica del messaggio) per ottenere una trasmissione efficiente dell'informazione su di un canale di comunicazione reale (con inevitabili limitazioni fisiche e presenza di disturbi).
Sostanzialmente, Shannon ha introdotto una definizione del concetto di informazione associata ad una grandezza chiamata *entropia*, riconoscendo le analogie formali e concettuali con quanto stabilito da Boltzmann a proposito dell'entropia termodinamica[1].

Nota 1: Il termine *entropia* deriva dal greco *"en"* (dentro) e *"tropé"* (rivolgimento), quindi "caos interno". L'entropia termodinamica è legata alla possibilità di trasformare l'energia termica in energia meccanica: se durante un processo l'entropia non varia, il processo è reversibile; se invece aumenta, l'energia disponibile diminuisce. Il concetto di entropia viene utilizzata anche in meccanica statistica; in questa disciplina un aumento di entropia corrisponde ad una diminuzione di ordine o, se vogliamo, ad una diminuzione della nostra conoscenza.

Modello del sistema di comunicazione

Un generico sistema di comunicazione può essere schematizzato nel seguente modello:



1. *Sorgente*: entità (persona o dispositivo) che genera messaggi;
2. *Trasmettitore*: dispositivo che trasforma i messaggi in segnali adatti ad essere trasmessi sul canale;
3. *Canale*: mezzo utilizzato per trasferire il segnale dal trasmettitore al ricevitore. In genere sul canale sono presenti disturbi (rumore) che possono corrompere il segnale trasmesso fino ad impedirne la corretta ricezione;
4. *Ricevitore*: dispositivo che elabora i segnali ricevuti per riconoscere i messaggi che possono essere stati realmente trasmessi;
5. *Destinatario*: entità (persona o dispositivo) al quale è indirizzato il messaggio trasmesso.

Misura dell'informazione

Si consideri una sorgente di messaggi (o simboli). La quantità di informazione trasmessa da un messaggio aumenta con l'aumentare dell'incertezza sul messaggio prodotto. Un messaggio scelto fra dieci messaggi possibili trasmette una quantità di informazione minore di quella trasmessa da un messaggio scelto fra un milione di messaggi possibili. La misura di questa incertezza è detta *entropia*. L'entropia è assunta come misura della quantità di informazione trasmessa con un messaggio da una sorgente. Tanto maggiore è la nostra conoscenza sul messaggio che sarà prodotto dalla sorgente, tanto minore sarà l'incertezza, minore l'entropia, e di conseguenza, minore l'informazione trasmessa.
La misura di informazione è legata all'incertezza associata all'emissione di ciascun simbolo (o messaggio): una grande incertezza sull'emissione di un simbolo equivale ad un grande contenuto informativo del simbolo stesso. Dunque, l'informazione associata ad un messaggio è legata alla sua probabilità.
Si consideri una sorgente S_X che può emettere un simbolo x_i alla volta con probabilità p_i su n simboli possibili. Ovvero:

$$S_X = \begin{Bmatrix} x_1 & x_2 & x_3 \dots & x_i & \dots.. & x_n \\ p_1 & p_2 & p_3 \dots & p_i & \dots.. & p_n \end{Bmatrix} \quad \text{dove} \quad \sum_{i=1}^{n} p_i = 1$$

Si definisce *autoinformazione* del messaggio x_i la quantità: $I(x_i) = \log_b(1/p_i) = - \log_b(p_i)$
Dove b rappresenta la base del logaritmo. La sua scelta determina l'unità di misura assegnata al contenuto di informazione: se la base è "*e*" l'unità di misura sarà il *nat*, se la base è "*10*" avremo l'*hartley*, mentre se la base è "*2*" parleremo di *bit* (acronimo di *Binary Digit*). In ingegneria, per questione di praticità, si utilizza la base binaria (b=2).

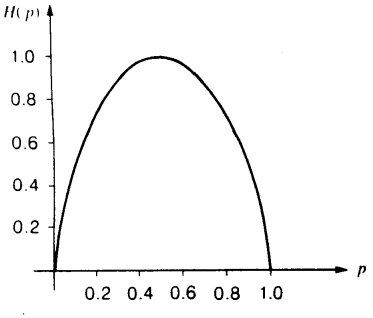
L'autoinformazione rappresenta la quantità di informazione associata al generico simbolo x_i.
Si definisce *entropia* della sorgente la quantità:

$$H(x) = \sum_{i=1}^{n} p_i I(x_i) = \sum_{i=1}^{n} p_i \log_2(\frac{1}{p_i}) \quad \text{[bit/simbolo]}$$

L'entropia della sorgente rappresenta il valor medio della quantità di informazione associata a ciascun simbolo.
Per una sorgente che può emettere *n* simboli possibili, l'entropia è massima quando tutti i simboli sono equiprobabili, ovvero quando $p_i = \frac{1}{n}$.
Pertanto risulta: $H_{max}(x) = \log_2 n$
Dunque: $0 \le H(x) \le \log_2 n$
Inoltre si definisce *entropia relativa* (o *efficienza di codifica*) della sorgente la quantità:
$H_r(x) = H(x) / H_{max}(x)$
E si definisce *ridondanza* della sorgente la quantità:
$R(x) = 1 - H_r(x)$ [bit/simbolo]
La ridondanza rappresenta la misura della rigidità della sorgente, ossia la misura del vincolo imposta dalla sorgente nella scelta dei simboli. Se tutti i simboli sono equiprobabili (massima scelta) la ridondanza è minima, ovvero R(x)=0.

Esempio n.1 *(Sorgente binaria)*
Data una sorgente che emette due soli simboli x_1 ed x_2, rispettivamente con probabilità p e 1-p calcolare l'entropia della sorgente.

$I(x_1) = \log_2(p)$
$I(x_2) = \log_2(1-p)$
$H(p) = p\log_2(p) + (1-p)\log_2(1-p)$



L'entropia è massima per p=1/2 (simboli equiprobabili) e vale $H_{max}(x) = 1$ [bit/simbolo].
L'entropia è minima quando uno dei due simboli ha probabilità massima (p=0 oppure p=1) e vale H(x) = 0 [bit/simbolo].

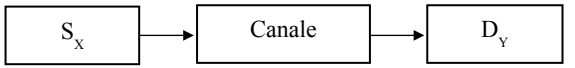
Esempio n.2
Data la seguente sorgente calcolare l'auoinformazione associata a ciascun simbolo, l'entropia della sorgente, l'entropia relativa e la ridondanza della sorgente.

$$S_X = \begin{Bmatrix} A & B & C & D \\ 0{,}2 & 0{,}5 & 0{,}2 & 0{,}1 \end{Bmatrix}$$

$I(A) = -\log_2(0{,}2) = 2{,}32$ [bit]
$I(B) = -\log_2(0{,}5) = 1$ [bit]
$I(C) = -\log_2(0{,}2) = 2{,}32$ [bit]
$I(D) = -\log_2(0{,}1) = 3{,}32$ [bit]

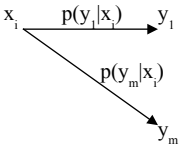
$H(x) = p_A I(A) + p_B I(B) + p_C I(C) + p_D I(D) = 0{,}2*2{,}32 + 0{,}5*1 + 0{,}2*2{,}32 + 0{,}1*3{,}32$
Dunque: H(x) = 1,76 [bit/simbolo]
Inoltre risulta: $H_{max}(x) = \log_2(4) = 2$ [bit/simbolo]
$H_r(x) = 1{,}76 / 2 = 0{,}88$
$R(x) = 1 - 0{,}88 = 0{,}12$ [bit/simbolo]

Trasmissione dell'informazione
Si consideri il seguente modello di sistema di trasmissione:



Dove S_X rappresenta una entità sorgente che emette simboli appartenenti ad un alfabeto $X = \{x_1\ x_2\ x_3\ \ldots..\ x_n\}$ (alfabeto della sorgente).
D_Y rappresenta una entità di destinazione in grado di ricevere simboli appartenenti ad un alfabeto $Y = \{y_1\ y_2\ y_3\ \ldots..\ y_m\}$ (alfabeto di destinazione).
Il canale di trasmissione si considera di tipo reale, dunque affetto da rumore, ovvero introduce dei disturbi durante la comunicazione. In questo caso il destinatario che riceve un simbolo y_i non è completamente certo che sia effettivamente stato trasmesso dalla sorgente il corrispondente simbolo x_i.
Il canale è dunque un dispositivo in grado di associare, in ogni istante, un simbolo y_i con il corrispondente simbolo x_i con probabilità $p(y_i|x_i)$.
Il canale può essere descritto in forma matriciale attraverso la *matrice di transizione* C:

$$C = \begin{bmatrix} p(y_1|x_1) & p(y_2|x_1) & \ldots.. & p(y_m|x_1) \\ p(y_1|x_2) & p(y_2|x_2) & \ldots.. & p(y_m|x_2) \\ \ldots & \ldots & \ldots & \ldots \\ p(y_1|x_n) & p(y_2|x_n) & \ldots.. & p(y_m|x_n) \end{bmatrix}$$



con $\sum_{j=1}^{m} p(y_j|x_i) = 1$ per ogni i = 1, 2, 3,… n

La probabilità $p(y_j|x_i)$ *(probabilità condizionata)* rappresenta la probabilità di ricevere il simbolo y_i dato che è stato trasmesso il simbolo x_i.

Osservazione: il canale ideale, canale indisturbato (senza rumore), è caratterizzato dalle probabilità $p(y_j|x_i) = 1$ per $j = i$ e $p(y_j|x_i) = 0$ per $j \neq i$; ovvero la matrice di transizione è una matrice diagonale con tutti gli elementi della diagonale principale di valore unitario.

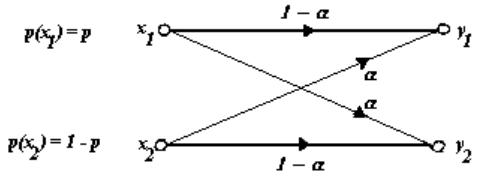
La probabilità che il destinatario riceva il generico simbolo y_j è:

$$p(y_j) = \sum_{i=1}^{n} p(x_i)\,p(y_j \mid x_i)$$

dove $p(x_i)$ è la probabilità di emissione del simbolo x_i e $p(y_j|x_i)$ la probabilità condizionata della coppia di simboli (x_i, y_j).

Esempio n.3
(Canale binario simmetrico)
Consideriamo una sorgente binaria che emette due simboli (x_1 e x_2) rispettivamente con probabilità p ed 1-p. Consideriamo un canale binario simmetrico caratterizzato dalla seguente matrice di transizione:

$$C = \begin{bmatrix} 1-\alpha & \alpha \\ \alpha & 1-\alpha \end{bmatrix}$$



Le probabilità di ricevere i simboli y_1 ed y_2 risultano:
$p(y_1) = p(1-\alpha) + (1-p)\alpha$
$p(y_2) = p\alpha + (1-p)(1-\alpha)$
Inoltre la probabilità totale di errore risulta: $p_e = p(x_1,y_2) + p(x_2,y_1)$
dunque $p_e = p(y_2|x_1)p(x_1) + p(y_1|x_2)p(x_2) = \alpha p + \alpha(1-p)$.

Osservazione: Se il canale fosse ideale ($\alpha=0$) risulterebbe: $p(y_1) = p$ ed $p(y_2) = 1-p$.

Capacità di trasmissione di un canale
Dato un sistema di comunicazione costituito da una sorgente, un canale ed un destinatario, si definiscono le seguenti grandezze:

1) Entropia della sorgente:

$$H(x) = \sum_{i=1}^{n} - p(x_i)\log_2 p(x_i)$$

L'entropia della sorgente rappresenta l'incertezza su quale simbolo verrà trasmesso.

2) Entropia del destinatario:

$$H(y) = \sum_{j=1}^{m} - p(y_j)\log_2 p(y_j)$$

L'entropia del destinatario rappresenta l'incertezza su quale simbolo sarà ricevuto.

3) Entropia congiunta:

$$H(x,y) = \sum_{i=1}^{n}\sum_{j=1}^{m} - p(x_i,y_j)\log_2 p(x_i,y_j)$$

L'entropia congiunta rappresenta l'incertezza che si ha quando si trasmette *x* e si riceve *y*.

4) Equivocazione:

$$H(x \mid y) = \sum_{j=1}^{m}\sum_{i=1}^{n} - p(y_j)p(x_i \mid y_j)\log_2 p(x_i \mid y_j)$$

L'entropia condizionata H(x|y) è detta equivocazione del canale, e rappresenta l'incertezza che sia stato trasmesso *x* quando si riceve *y*.
L'equivocazione è dunque l'incertezza del destinatario riguardo a ciò che è stato effettivamente trasmesso. L'equivocazione è la misura dell'informazione persa durante il processo di trasmissione.

5) Entropia di rumore:

$$H(y \mid x) = \sum_{i=1}^{n}\sum_{j=1}^{m} - p(x_i)p(y_j \mid x_i)\log_2 p(y_j \mid x_i)$$

L'entropia condizionata H(y|x) è detta entropia di rumore, o anche *irrilevanza*, e rappresenta l'incertezza di ricevere *y* quando è stato trasmesso *x*.
L'entropia di rumore è dunque l'incertezza del mittente rispetto a ciò che sarà ricevuto.

Osservazione:
Dalla teoria delle probabilità risulta: $p(x_i,y_j) = p(y_j|x_i)p(x_i) = p(x_i|y_j)p(y_j)$

Tra le diverse entropie dichiarate esiste la seguente relazione:
$H(x,y) = H(x) + H(y|x) = H(y) + H(x|y)$

Osservazione:
Quando il canale è indisturbato (senza rumore) risulta:
$H(y|x) = H(x|y) = 0$ e dunque $H(y) = H(x)$
Ovvero il mittente conosce esattamente ciò che sarà ricevuto, ed il destinatario conosce ciò che è stato trasmesso.

Si definisce *ritmo di trasmissione* o *informazione mutua* la seguente grandezza:
$I(x,y) = H(x) - H(x|y) = H(y) - H(y|x)$
Ossia $I(x,y) = H(x) + H(y) - H(x,y)$

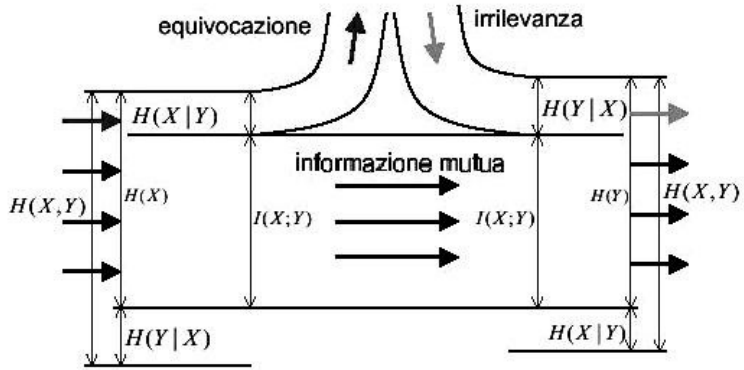
L'informazione mutua rappresenta la quantità di informazione che attraverso il canale passa dalla sorgente al destinatario. Essa dipende sia dalla sorgente sia dal canale.
Osservazione: Per un canale indisturbato risulta: $I(x,y) = H(x) = H(y)$

Si definisce *capacità del canale* la massima informazione mutua che può attraversare un dato canale, ovvero il massimo ritmo di trasmissione su quel canale.
La capacità di canale si indica nel seguente modo:
$C = I_{max}(x,y)$

Osservazione: Il calcolo della capacità di un canale è un compito molto difficile. Alcuni canali presentano problemi di calcolo pressoché insolubili.

Esempio n.4
(Capacità di un canale binario simmetrico)
Si consideri un canale binario simmetrico, come nell'esempio n.3.
Il massimo ritmo di informazione, la capacità del canale, sarà ottenuto quando i simboli della sorgente sono equiprobabili, ossia quando $p(x_1) = p(x_2) = ½$.
Allora risulta $H(x) = 1$ [bit/simbolo] ed $H(x|y) = -(1-\alpha)\log_2(1-\alpha) - \alpha\log_2(\alpha)$ [bit/simbolo].
Dunque $C = H(x) – H(x|y) = 1 + (1-\alpha)\log_2(1-\alpha) + \alpha\log_2(\alpha)$ [bit/simbolo].

Meccanismo di trasferimento dell'informazione
Data una sorgente, un canale ed un destinatario, il meccanismo di trasferimento dell'informazione, dalla sorgente al destinatario, può essere schematizzato attraverso il seguente diagramma (*diagramma di Berger*):



Al lato della sorgente: Una parte dell'informazione H(x) fornita dalla sorgente, l'equivocazione H(x|y), viene persa e non può essere convogliata verso il destinatario. L'equivocazione rappresenta la parte non utilizzabile dell'informazione fornita dalla sorgente al canale; essa è l'equivoco che resta su *x* se conosciuto attraverso *y*.

Al lato del destinatario: A formare l'informazione H(y) fornita al destinatario concorre una parte indicata col nome di irrilevanza H(y|x). Questo termine è irrilevante ai fini dello scambio di informazione fra le parti. L'irrilevanza rappresenta la frazione di informazione, disponibile all'uscita del canale, non proveniente direttamente dalla sorgente e irrilevante ai fini della conoscenza di *x* attraverso *y*.

Codifica dell'informazione
I simboli emessi dalla sorgente subiscono varie trasformazioni (operazioni di *codifica*), prima di essere trasmessi attraverso un canale di trasmissione. Le operazioni di codifica consistono nell'associare ai simboli o a gruppi di simboli (*parole*) emessi ed appartenenti all'alfabeto di sorgente, simboli o sequenze di simboli estratti da un altro alfabeto, diverso in genere dal precedente. Il dispositivo che opera le trasformazioni suddette prende i nome di *codificatore*, mentre le trasformazioni stesse prendono il nome di *codice*.
Riguardo alle operazioni di codifica si distinguono tre tipi di codici: codici di sorgente, codici di canale e codici di linea.
La *codifica di sorgente* ha lo scopo di ridurre la lunghezza dei messaggi da trasmettere, senza per questo perdere informazione, garantendo la possibilità di univocità di decodifica da parte del destinatario. Tali trasformazioni sono individuate a partire dalla conoscenza delle caratteristiche della sorgente.
La *codifica di canale* ha lo scopo di controllare e ridurre la presenza di errori dovuti ai disturbi introdotti dal canale. Tale obiettivo è raggiunto mediante l'introduzione di una opportuna ridondanza, che consente la rilevazione o addirittura la correzione degli errori.
La *codifica di linea* ha come obiettivo di adattare al meglio le caratteristiche dei segnali associati ai simboli trasmessi a quelle dei mezzi trasmissivi impiegati.

Teorema fondamentale di Shannon
Data una sorgente con entropia *H* ed un canale di capacità *C* risulta:
- se $H \leq C$ allora è possibile codificare i simboli della sorgente in modo da essere trasferiti sul canale con una frequenza di errore (equivocazione) arbitrariamente piccola, ossia $H(x|y) < \varepsilon, \ \forall\varepsilon>0$;
- se $H > C$ allora non è possibile, in alcun modo, codificare i simboli della sorgente in modo da avere una frequenza di errore arbitrariamente piccola, ossia $H(x|y) < H – C + \varepsilon, \ \forall\varepsilon>0$.

Osservazione: Il teorema di Shannon non dice come individuare o tantomeno come costruire un codice di canale per ridurre l'equivocazione; a tuttoggi non è stata individuata nessuna regola per costruire un codice di canale che consenta di avvicinare i limiti teorici stabiliti dal teorema.

Conclusioni

La teoria dell'informazione, così come sviluppata da Shannon, affronta il problema di realizzare una comunicazione efficiente ed esente da errori, attraverso un canale di trasmissione reale (cioè affetto da rumore). L'intero problema si riduce essenzialmente ad affrontare due attività: eliminare dai messaggi la ridondanza superflua *(codifica di sorgente)* e poi aggiungervi quel tipo di ridondanza (scelta ad hoc) che consente la rilevazione e la correzione degli errori che si presentano in trasmissione *(codifica di canale)*.

Il lavoro di Shannon affronta il problema della comunicazione efficiente in modo molto generale, illustrando il modo di procedere in linea teorica; ma esistono grandi difficoltà matematiche per trattare i canali complessi.