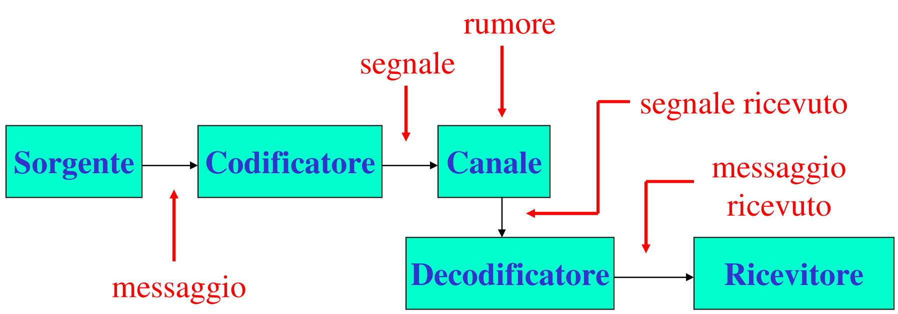
Entropia ed informazione

Primi elementi sulla teoria della misura dell'informazione

Per trasmettere l'informazione è necessaria una
rete di comunicazione,
che, secondo l'approccio teorico di Claude E. Shannon e Warren Weaver
(ingegneri presso i Bell Laboratories, USA, 1949),
si può così schematizzare:



Nel trasmettere l'informazione occorre adottare un
criterio di efficienza:
minima distorsione, massima velocità, minimo costo.

Per tutti i sistemi di comunicazione il **problema fondamentale** consiste nello **stabilire un'adeguata misura dell'informazione**, e nel servirsene per migliorare la loro efficienza.

Cerchiamo di definire quanta informazione è contenuta nell'evento

$$\{X = x\}$$ X variabile aleatoria

Prima di tutto, è importante capire che in questo contesto il termine informazione non riguarda il contenuto semantico di un messaggio.

Informazione qui significa semplicemente una **misura della libertà di scelta di cui si dispone nello scegliere un messaggio dall'insieme di quelli disponibili**, molti dei quali potranno anche essere privi di significato.

La quantità di informazione ***I*** portata dal messaggio $\{X = x\}$ si suppone ragionevolmente che:

1) dipenda dalla probabilità *$1 \geq p \geq 0$* che *X* sia uguale a *x:*

$$I=I(p)$$

2) sia tanto più grande quanto più rara è la sua occorrenza, cioè, in altre parole, *I* è decrescente in ***p***

3) sia una grandezza non negativa: $I(p) \geq 0$

4) sia additiva rispetto alla somma di messaggi. In altri termini, è naturale supporre che, se *X* e *Y* sono 2 v.a. indipendenti:

$$P\{X = x\} = p, P\{Y = y\} = q, P\{X = x, Y = y\} = pq$$

allora deve verificarsi che

$$I(pq) = I(p) + I(q)$$

Dalle 1), 2) e 3) siamo indotti ad ipotizzare una funzione del tipo

$$I \propto \frac{1}{p}$$

Ma dalla 4) si può dedurre subito che per trasformare prodotti in somme occorre ricorrere ai logaritmi; di conseguenza:

$$I \propto \log\frac{1}{p} = -\log p$$

Per quanto riguarda la scelta della base del logaritmo, in generale questa è arbitraria; tuttavia, si preferisce usare 2, con riferimento al sistema binario usato nei calcolatori elettronici.

In questo modo, l'informazione viene misurata in *bit*, ovvero in cifre binarie (*binary digits*)

In definitiva, possiamo affermare che l'informazione ***I*** (parziale) contenuta in un singolo messaggio $\{X = x\}$ di probabilità ***p*** vale:

$$I(p) = -\log_2 p$$

Consideriamo adesso una v.a. *X*, che possa assumere i valori

$x_1, x_2, x_3, \ldots\ldots, x_n$ con probabilità $p_1, p_2, p_3, \ldots\ldots, p_n$

Si definisce **valore atteso dell'informazione contenuta in *X***

oppure

informazione media contenuta in tutti i valori possibili di *X*

oppure

quantità media di informazione (o di *sorpresa*) che riceviamo quando osserviamo il valore che *X* ha assunto oppure

misura d'incertezza a priori di *X*

$$H(X) = -\sum_{i=1}^{n} p_i \log_2 p_i \Longleftrightarrow H(X) = E\left[-\log_2 p(X)\right]$$

La grandezza *H(X)* è chiamata entropia (di Shannon) di *X*

Notiamo tre proprietà dell'entropia *H(X)*:

•l'entropia di un "bit casuale" è uguale a 1

•l'entropia è massima quando tutti i valori x_i (i=1,2,…,n) che può assumere la v.a. *X* sono equiprobabili:

$$p_i = p = 1/n$$

In questo caso si ha che:

$$H(X)_{MAX} = \log_2 n$$

•Se si conosce a priori il messaggio (cioè, il valore che assume la v.a. *X*), allora ***H(X) = 0***

In statistica ***l'entropia di Shannon*** è utilizzata come ***indice di eterogeneità*** di un insieme di dati, di frequenza nota, e spesso viene fornita in forma normalizzata, cioè divisa per l'entropia massima.

Verifichiamo la prima proprietà dell'entropia *H(X)*:
•l'entropia di un "bit casuale" è uguale a 1:

x_i	p_i
0	1/2
1	1/2

$$H(X) = -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2}$$

$$H(X) = -\log_2\frac{1}{2} = -(-1) = +1$$

Verifichiamo una parte della seconda proprietà dell'entropia *H(X)*:
•se i valori che può assumere la v.a. *X* sono equiprobabili, allora:

x_i	p_i
x_1	$1/m$
x_2	$1/m$
$\vdots$	$\vdots$
x_m	$1/m$

$$H(X) = -\sum_{i=1}^{m} p_i \log_2 p_i$$

$$H(X) = -m\frac{1}{m}\log_2\frac{1}{m} = -\log_2\frac{1}{m}$$

$$H(X) = -\log_2 1 + \log_2 m = \log_2 m$$

SI DIMOSTRA CHE $= H(X)_{MAX}$

Esempio 1			
Messaggio	$\{X = x_1\}$	$\{X=x_2\}$	$\{X=x_3\}$
Probabilità del messaggio	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{6}$
Contenuto di informazione del messaggio	$-\log_2 \frac{1}{2} = 1$	$-\log_2 \frac{1}{3} = 1.58$	$-\log_2 \frac{1}{6} = 2.58$

Quanto vale ***H(X)*** $\equiv$ misura dell'informazione media dell'insieme dei messaggi nel suo complesso?

$$H(X) = \frac{1}{2}(1) + \frac{1}{3}(1.58) + \frac{1}{6}(2.58) = 1.46 \text{ bit}$$

Esempio 2

Trovare l'**entropia massima** di un messaggio costituito da 5 lettere, essendo il numero totale di lettere nell'alfabeto uguali a 32.

Il numero di messaggi possibili composti da 5 lettere, anche tutte uguali, è dato dal numero di disposizioni con ripetizione di 32 lettere a gruppi di 5:

$$D^r_{32,5} = 32^5$$

Dato che si suppone che ogni messaggio abbia la stessa probabilità di verificarsi (messaggi equiprobabili), allora **l'entropia massima è**:

$$H(X)_{MAX} = \log_2 32^5 = 5\log_2 32 = 5 \times 5 = 25 \text{ bit}$$

Si noti che, essendo a priori tutti i messaggi equiprobabili, l'informazione parziale contenuta in un singolo messaggio è uguale all'informazione media:

$$I_{x_i} = -\log_2 p_{x_i} = -\log_2 \frac{1}{32^5} = \log_2 32^5 = 25 \text{ bit}$$

Ridondanza

L'alfabeto, nel caso della lingua inglese, contiene 26 lettere più l'intervallo fra parole contigue, cioè 27 simboli.

Allora, il repertorio di *messaggi* più elementare di tale lingua è costituito da 27 simboli.

Se si assume che la probabilità di ogni messaggio (simbolo) sia la stessa,l'entropia è data da:

$$H_{massima} = \log_2 27 = 4.76\ bit$$

In altre parole, sono necessari circa 5 bit per codificare ogni simbolo:

00000, 00001, 00010, 00100, 00111, ……

Tuttavia, l'effettiva quantità di informazione trasmessa da ogni lettera è in realtà minore, perché le diverse lettere non compaiono con uguale frequenza. Ad esempio,

p(a)=0.06, p(b)=0.01, p(c)=0.02, p(d)=0.03, p(e)=0.1, …..

Allora, l'entropia diventa:

$$H = -\sum_{i=1}^{27} p_i \log_2 p_i = 4\ bit$$

Ancora una volta, ci si rende conto che l'effettiva quantità di informazione trasmessa da ogni lettera è molto più bassa, a causa di un'importante proprietà comune a tutte le lingue e chiamata **ridondanza**. Ad esempio,

- **la probabilità di q-u è praticamente 1**
- **la probabilità di j-x è praticamente 0**
- **la probabilità di tio-n è molto elevata**
- **la probabilità di oug-k è minima**

Attraverso una serie di esperimenti si arriva a stabilire che l'effettiva quantità di informazione trasmessa da ogni lettera è

$$H_{effettiva} \approx 2\ bit$$

Possiamo quindi assumere come misura della libertà relativa di scelta dei simboli rispetto alla massima possibile il seguente rapporto, detto anche **entropia relativa** del messaggio:

$$\frac{H_{effettiva}}{H_{massima}} = \frac{2}{4.76} = 0.42$$

Adesso possiamo definire **la ridondanza**:

$$ridondanza = 1 - \frac{H_{effettiva}}{H_{massima}} = 0.58$$

In definitiva, la **ridondanza rappresenta la percentuale di informazione del messaggio che rimane automaticamente definita dai vincoli statistici della lingua**.