Informazione ed entropia

- In generale osservando un flusso di dati si può tentare di svelare il contenuto originale. Un problema connesso a questo è la misura dell'informazione estraibile da (o contenuta in) un flusso di dati.
- Il concetto di Entropia informazionale nasce proprio dall'esigenza di quantificare il contenuto informativo di sequenze astratte di segnali.
- Affinché il concetto risponda bene alle nostre esigenze deve dare un contenuto informativo nullo alle sequenze di numeri casuali.

Contenuto informativo di un evento

- Misurare la rarità di un evento è il punto di partenza per comprendere il legame tra informazione e disordine.
- Una funzione che misuri il contenuto informativo di un evento A deve dipendere dalla sua probabilità.
- Ma non può essere p(A) perché più è piccola la probabilità e più sappiamo sul sistema.
- Se si verificano due eventi indipendenti l'informazione acquisita deve essere la somma delle informazioni legate ai singoli eventi:
- $I(A \cap B) = I(A) + I(B)$;che, essendo $P(A \cap B) = P(A)P(B)$, equivale alla proprietà di fattorizzazione della f : $f(x \cdot y) = f(x) + f(y)$:

Contenuto di informazione secondo Shannon

La funzione $\mathcal{I}(x)$ misura la **quantità d'*I*nformazione** contenuta nell'elemento x dell'insieme $\mathcal{A}$.

La funzione $\mathcal{I}(x)$ soddisfa le proprietà della funzione logaritmo.

Gli eventi certi non danno informazione ($\mathcal{I}=0$)

Gli eventi poco probabili danno molta informazione

Shannon definisce il Contenuto di informazione come: $\mathcal{I}(X=x) = \log_2\left(\frac{1}{p(x)}\right)$ misurato in bit.

Shannon definisce il Contenuto di informazione come: $\mathcal{I}$(X=x)= $\log_2\left(\frac{1}{p(x)}\right)$ misurato in bit.

Esempi

Quanto vale $\mathcal{I}$(x) nei casi

- $\mathcal{A}$=$\{\mathrm{T},\mathrm{C}\}$ moneta non truccata, le probabilità sono $1/2,1/2$.

Calcoli: $\mathcal{I}(\mathrm{T})$……

$\mathcal{I}(\mathrm{C})$……

- $\mathcal{A}$= $\{\mathrm{T},\mathrm{C}\}$ moneta **truccata**, le probabilità sono $1/3,2/3$

Calcoli: $\mathcal{I}(\mathrm{T})$……

$\mathcal{I}(\mathrm{C})$……

- $\mathcal{A}$= $\{00,01,10,11\}$ **stringhe binarie di lunghezza $\mathcal{L}$=2**, con probabilità tutte uguali a $1/4$.

Calcoli: $\mathcal{I}(X=00)$ ……

$\mathcal{I}(X=01)$ ……

. $\mathcal{I}(X=10)$ …… .

. $\mathcal{I}(X=11)$ ……

- $\mathcal{A}^2$= $\{00,01,10,11\}$ **stringhe binarie di lunghezza $\mathcal{L}$=2**, con probabilità $\left\{\frac{1}{3},\frac{1}{6},\frac{1}{6},\frac{1}{3}\right\}$

Calcoli: $\mathcal{I}(X=00)$ ……

$\mathcal{I}(X=01)$ ……

. $\mathcal{I}(X=10)$ …… .

. $\mathcal{I}(X=11)$ ……

Lettere e microstati, stringhe e macrostati

- $\mathcal{A}$ è un insieme di lettere. Un insieme (finito) di lettere forma una stringa.
- È utile introdurre una funzione stato che sceglie un elemento in $\mathcal{A}$.

Le possibili lettere di un alfabeto costituiscono i possibili microstati.

- Una singola stringa è formata da tante lettere quindi è formata da tanti microstati.
- Considero l'insieme $\mathcal{A}^L$ di tutte le stringhe di lunghezza **L. I microstati adesso sono tutte le possibili stringhe.** Questo insieme può partizionato ed ogni partizione sarà formata da gruppi di stringhe diverse (ma equivalenti); ogni elemento della partizione è chiamato macrostato.

  - $\mathcal{A}=\{\mathrm{T},\mathrm{C}\}$, i microstati sono T e C.
  - $\mathcal{A}^2=\{\mathrm{TT},\mathrm{TC},\mathrm{CT},\mathrm{CC}\}$, i microstati sono……

Possiamo fare delle partizioni di $\mathcal{A}^2$.

- Esempio 1

microstato≡macrostato	TT	TC	CT	CC
probabilità	1/4	1/4	1/4	1/4

- Esempio 2

Macrostato	TT	{TC, CT}	CC
probabilità	1/4	2/4	1/4

- Esempio 3

Macrostato	TT	{TC, CT}	CC
probabilità	1/3	1/3	1/3

Entropia secondo Shannon

- Il termine Entropia assume significati molteplici. Noi ci occuperemo dell'entropia informazionale ed utilizzeremo l'approccio classico che si deve a Claude Shannon.
- Per i suoi contributi negli anni dopo la seconda guerra mondiale, viene considerato il "padre della teoria dell'informazione".

- Il termine entropia nasce in fisica ad opera di **Rudolf Clausius** (a metà del 1800) come funzione di stato che misura il livello di disordine dei sistemi termodinamici e contribuisce all'equilibrio termodinamico.
- L'etimologia del termine viene da **εν**, "dentro", e **τροπή** "movimento" e significa "cambiamento interno". Un cambiamento del sistema non legato alla sua energia.
- In meccanica statistica ed in teoria dei campi (in particolare nella formulazione di **Gibbs**) si lega indissolubilmente l'entropia termodinamica al numero degli stati microscopici che caratterizzano uno stato macroscopico. Precisamente l'entropia è il logaritmo naturale del numero di tali stati.
- L'entropia di una sorgente (di **Shannon**) è l'applicazione naturale dei concetti fondamentali della meccanica statistica ai sistemi informazionali.

L'entropia di una sorgente è definita come il valore atteso dell'autoinformazione, ovvero l'informazione media contenuta in ogni messaggio emesso.

Equivalentemente si può definire l'entropia Shannon di un insieme $X = (x, \mathrm{A}_x, P_x)$ come l'informazione media di un evento.

Sempre fornendo la misura in bit si può scrivere la formula

$$H(X) = \sum_{x \in A_x} P(x) \cdot \Im(\mathrm{x}) = \sum_{x \in A_x} P(x) \log_2 \frac{1}{P(x)}$$

Oppure $\mathcal{I}(\mathrm{x})$

$$H(X) = \sum_{i=1}^{k} p_i(x) \log_2 \frac{1}{p_i(x)}$$

http://www.fmboschetto.it/didattica/entropologia/entropia_informazionale.html

http://www.fmboschetto.it/didattica/entropologia/fisica_statistica.html

Entropia secondo Shannon
Alcuni esempi

$H(X) = \sum_{x \in A_x} P(x) \log_2 \frac{1}{P(x)}$ $\qquad$ $H(X) = \sum_{i=1}^{k} p_i(x) \log_2 \frac{1}{p_i(x)}$

Calcolo dell'entropia di Shannon per un insieme X

- Con alfabeto $\mathcal{A}_x$={T, C }, con microstati aventi probabilità $\left\{\frac{1}{2}, \frac{1}{2}\right\}$
- Con alfabeto $\mathcal{A}_x$={T, C }, con microstati aventi probabilità $\left\{\frac{1}{3}, \frac{2}{3}\right\}$
- Con alfabeto $\mathcal{A}_x$=$\mathcal{A}^2$ = {TT, TC, CT, CC }, con microstati aventi probabilità $\left\{\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right\}$
- Con alfabeto $\mathcal{A}_x$= $\mathcal{A}^2$ = {TT, TC, CT, CC }, con microstati aventi probabilità $\left\{\frac{1}{3}, \frac{1}{6}, \frac{1}{6}, \frac{1}{3}\right\}$

Calcoliamo quanto vale $H(X) = \sum_{i=1}^{k} p_i(x) \log_2 \frac{1}{p_i(x)}$ nei diversi casi proposti

- $\mathcal{A}_x$={T, C } moneta non truccata, le probabilità sono 1/2,1/2.

Calcoli: $H(X) = \sum_{i=1}^{k} p_i(x) \log_2 \frac{1}{p_i(x)}$ = ……

……

- $\mathcal{A}_x$= {T, C } moneta **truccata**, le probabilità sono 1/3,2/3

Calcoli: $H(X) = \sum_{i=1}^{k} p_i(x) \log_2 \frac{1}{p_i(x)}$ = ……

……

- $\mathcal{A}_x$=$\mathcal{A}^2$= {00,01, 10, 11 } **stringhe binarie di lunghezza $\mathcal{L}$=2**, con probabilità tutte uguali a 1/4.

Calcoli: $H(X) = \sum_{i=1}^{k} p_i(x) \log_2 \frac{1}{p_i(x)}$ = ……

……

- $\mathcal{A}_x$=$\mathcal{A}^2$= {00,01, 10, 11 } **stringhe binarie di lunghezza $\mathcal{L}$=2**, con probabilità $\left\{\frac{1}{3}, \frac{1}{6}, \frac{1}{6}, \frac{1}{3}\right\}$

Calcoli: $H(X) = \sum_{i=1}^{k} p_i(x) \log_2 \frac{1}{p_i(x)}$ = ……

……

Possiamo anche fare delle partizioni di $\mathcal{A}^2$ per ritrovare le precedenti configurazioni.

- Esempio 1

Macrostato	TT	TC	CT	CC
probabilità	1/4	1/4	1/4	1/4

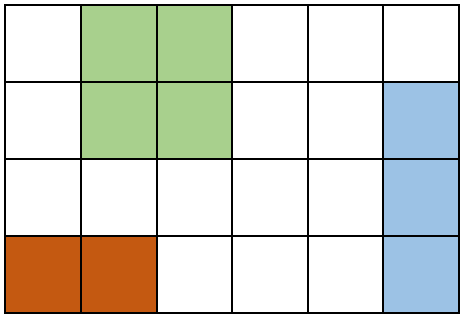
- Esempio 2

Macrostato	TT	{TC, CT}	CC
probabilità	1/4	2/4	1/4

- Esempio 3

Macrostato	TT	{TC, CT}	CC
probabilità	1/3	1/3	1/3

SCHEDA: la fisica della tua camera



La tua stanza può essere schematizzata con un quadrato di 24 quadretti.
I 4 quadretti verdi indicano la scrivania, i 2 quadretti marroni sono la libreria e i 3 azzurri, l'armadio.
La tua stanza come sistema fisico può essere descritto da due stati: S_1 = ordinata e S_2 = disordinata. Immagina di avere una penna P, un libro L ed una camicia C.

Usando la scheda disponi P; L e C in modo ....ragionevolmente **ordinato S_1**... 😊
Quante possibili posizioni puoi scegliere per ciascun oggetto mantenendo la stanza ordinata?

Usando la scheda puoi anche disporre P; L e C in modo ....**disordinato S_2**... ☹ Quante possibili posizioni puoi scegliere per ciascun oggetto mantenendo la stanza disordinata?
Sei autorizzato ad essere veramente disordinato quindi potresti, per esempio, mettere la penna e la camicia nello stesso riquadro!!!

Gli stati S_1 e S_2 della stanza sono due esempi di stati macroscopici.
Gli stati invece di L, P e C sono invece detti stati microscopici.

Codifica di un messaggio

In un testo scritto in inglese, è vero che tutte le lettere capitano con la stessa frequenza? Cerca di interpretare la seguente tabella.

i	a_i	p_i	$h(p_i)$
1	a	.0575	4.1
2	b	.0128	6.3
3	c	.0263	5.2
4	d	.0285	5.1
5	e	.0913	3.5
6	f	.0173	5.9
7	g	.0133	6.2
8	h	.0313	5.0
9	i	.0599	4.1
10	j	.0006	10.7
11	k	.0084	6.9
12	l	.0335	4.9
13	m	.0235	5.4
14	n	.0596	4.1
15	o	.0689	3.9
16	p	.0192	5.7
17	q	.0008	10.3
18	r	.0508	4.3
19	s	.0567	4.1
20	t	.0706	3.8
21	u	.0334	4.9
22	v	.0069	7.2
23	w	.0119	6.4
24	x	.0073	7.1
25	y	.0164	5.9
26	z	.0007	10.4
27	-	.1928	2.4
	$\sum_i p_i \log_2 \frac{1}{p_i}$		4.1

Table 2.9. Shannon information contents of the outcomes a–z.

COMPLETARE

- Cosa rappresenta il ventisettesimo simbolo? ……..
- Cosa rappresenta l'ultima colonna?

?

- Calcola il contenuto di informazione, espresso in bit, dell'evento x=**z** 10,48
- Calcola il contenuto di informazione, espresso in bit, dell'evento x=**e** ? 3,45
- Commenta i due risultati precedenti
  ……
  ……
  ……
- Verifica che l'entropia di una lettera estratta a caso da un testo in inglese è circa 4,11 se usiamo i dati delle probabilità riportati in tabella.
- Quanto varrebbe l'entropia se i simboli fossero tutti equiprobabili?
  ……
- Si può dimostrare che l'entropia viene massimizzata da distribuzioni di probabilità uniformi.
- Si può dimostrare che l'entropia è additiva per variabili stocastiche indipendenti.

Quante sono le lettere ed i simboli che ci permettono di scrivere in inglese?
Se tu volessi usare $\{0,1\}$per codificare tali simboli, quanto dovranno essere lunghe le stringhe di 0 e 1?

L'approccio probabilistico

Una delle domande fondamentali quando si è di fronte ad un flusso o un insieme di dati riguarda la loro regolarità.
L'attaccante si chiede se può estrarre elementi conoscitivi dalla sequenza dei dati.
Il **difensore** si chiede se stia rendendo pubblica qualche parte dell'informazione che intende custodire o trasmettere.
In entrambi i casi l'approccio probabilistico consente di valutare **quantitativamente** il problema.

Attesa probabilistica

Data una variabile stocastica discreta η che può assumere i valori $x^1; x^2; : : : ; x^M$, si definisce valore d'attesa (o semplicemente attesa) di una funzione di η la sua somma sui valori x^i pesata con le probabilità degli stessi: $E[f(\eta)] := \sum_{i=1}^{M} f(x^i) \cdot p_i$.

Se la cardinalità M è finita la somma esiste sempre; se la variabile stocastica assume un insieme di valori numerabile, la somma si estende all'infinito e non sempre converge. Le funzioni limitate (funzioni di prova) sono sempre sommabili.

La frequenza dei caratteri

Una delle caratteristiche stocastiche di una sorgente è la frequenza dei caratteri (sia in senso stretto, quando sono lettere, sia in senso generalizzato come sequenza di 8 bit (numeri tra 0 e 255).
Un attaccante che osservi un lungo flusso di caratteri generato da una sorgente può innanzi tutto calcolare le frequenze dei caratteri ed osservarne la eventuale stabilità.
Questa operazione si chiama analisi delle frequenze e consente di decrittare tutti i cifrari sostituzionali (se sufficientemente lunghi).
Calcolando la frequenza di un carattere nel crittogramma e comparandola con quelle dei caratteri nella lingua in cui scritto il messaggio, possiamo dedurre quale carattere sostituisca.

Osservazioni

- In ogni linguaggio le lettere dell'alfabeto vengono utilizzate con una certa frequenza. Ogni lettera ha una frequenza diversa.
- Nella lingua parlata o scritta l'ipotesi di **indipendenza dei caratteri non è valida**, ma la legge dei grandi numeri sussiste.
- Questa proprietà suggerisce un metodo per la crittanalisi dei testi codificati con cifrature sostituzionali:
  Si calcolano le frequenze dei caratteri cifrati e si identificano con le lettere dell'alfabeto che posseggono la stessa frequenza.

Più i testi sono lunghi e meno è possibile commettere degli errori nelle attribuzioni.
Ovviamente questo metodo è velocissimo perché scala linearmente con la lunghezza del messaggio.
In generale: più lungo è il messaggio cifrato, più facile è la decrittazione.