Fondamenti di Informatica

per la Sicurezza

a.a. 2008/09

Teoria dell'Informazione

Stefano Ferrari

UNIVERSITÀ DEGLI STUDI DI MILANO
DIPARTIMENTO DI TECNOLOGIE DELL'INFORMAZIONE

Informazione

Il termine "informazione" è usato in molti contesti e con molte accezioni.

Portano informazione:

- una certa distribuzione di gocce di inchiostro su un foglio;
- una certa sequenza di lettere dell'alfabeto;
- una certa sequenza di parole della lingua italiana.

Informazione

Si individuano diversi livelli di informazione:

- sintattico:
  - l'informazione sintattica è legata alle configurazioni del supporto fisico;
- semantico:
  - l'informazione semantica è legata al significato attribuibile alle diverse configurazioni del supporto fisico;
- pragmatico:
  - l'informazione pragmatica è legata al valore attribuibile alle diverse configurazioni del supporto fisico.

Teoria dell'informazione

La branca dell'informatica nota come
teoria dell'informazione
studia l'informazione di tipo sintattico.

Misura dell'informazione (1)

Qual'è la natura dell'informazione?

Come si può misurare l'informazione?

Per rispondere a queste domande proviamo ad analizzare alcuni casi.

Misura dell'informazione (2)

Poniamo il caso di dover comunicare se un determinato evento è accaduto oppure no.

Consideriamo le seguenti modalità per comunicare l'evento:

1. organizziamo un falò da qualche decina di metri cubi di legna;
2. accendiamo un cerino.

Quale modalità trasferisce più informazione?

Misura dell'informazione (3)

La risposta è chiara: sia chi vede il cerino, sia chi vede il falò hanno la stessa informazione.

Quindi, si può trarre una prima conclusione:

la quantità di informazione dipende dall'evento, non dal mezzo di comunicazione!

Misura dell'informazione (4)

Consideriamo ora le seguenti domande:

a) Quanti sono i sette nani?

b) Quale lato della moneta che ho appena lanciato è uscito?

Quale risposta fornisce più informazione?

Misura dell'informazione (5)

La risposta alla domanda a) è conosciuta.

La risposta alla domanda b), magari non è interessante, ma toglie un dubbio.

Possiamo trarre una seconda conclusione:

l'informazione è legata all'incertezza.

Misura dell'informazione

Ipotizziamo un messaggio che contenga un simbolo x scelto da un insieme $X = \{x_i \mid 1 \leq i \leq n\}$ di n simboli (detto alfabeto).

Attraverso una funzione $I(\cdot)$, vorremmo misurare l'informazione contenuta nel messaggio, $I(x)$.

Quali proprietà deve avere la funzione $I(\cdot)$?

Proprietà dell'informazione (1)

L'informazione portata da una sequenza di simboli deve essere la somma dell'informazione portata dai singoli simboli che compongono la sequenza stessa:

$$I(x_i x_j) = I(x_i) + I(x_j)$$

Proprietà dell'informazione (2)

Se il simbolo x_i è meno frequente (o meno probabile) del simbolo x_j, l'informazione portata da x_i deve essere maggiore dell'informazione portata da x_j:

$$p(x_i) \leq p(x_j) \Rightarrow I(x_i) \geq I(x_j)$$

Proprietà dell'informazione (3)

Se due simboli sono equiprobabili, l'informazione da essi portata deve essere la stessa:

$$p(x_i) = p(x_j) \Rightarrow I(x_i) = I(x_j)$$

Proprietà dell'informazione (4)

Se è certo che un dato simbolo apparirà sul messaggio, allora l'informazione portata dal messaggio è nulla:

$$p(x_i) = 1 \Rightarrow I(x_i) = 0$$

Proprietà dell'informazione (5)

Meno probabile è un simbolo, maggiore è l'informazione da esso portata:

$$p(x_i) \rightarrow 0 \Rightarrow I(x_i) \rightarrow \infty$$

Funzione informazione

Una funzione che gode delle precedenti proprietà è:

$$I(x) = -\log_2 p(x),\ p(x) > 0$$

Questa funzione ha il vantaggio di valere 1 se l'insieme di simboli che il messaggio può contenere è costituito da soli due simboli equiprobabili.

In tal caso, ogni messaggio porta 1 bit!

Bit e informazione (1)

A questo punto ci sono due significati per il termine bit:

- unità di misura della capacità (o dell'ingombro) di una rappresentazione binaria;
- unità di misura dell'informazione.

Bit e informazione (2)

Per codificare 256 simboli equiprobabili, si usano 8 cifre binarie.

Ogni cifra binaria porta 1 bit di informazione.

Se i simboli non fossero equiprobabili, alcune cifre potrebbero portare (in media) più di 1 bit, e altre meno di 1 bit.

Indovinare un numero (1)

Esempio

Regole del gioco "Indovina il numero":

1. una persona pensa un numero tra 1 e 128;
2. per scoprire tale numero gli si possono fare delle domande;
3. a tali domande, la persona può rispondere solo "Sì" o "No".

Qual è il numero minimo di domande necessarie per indovinare il numero nascosto?

Indovinare un numero (2)

Risposta: 7.

Traccia:

- la conoscenza del numero porta 7 bit di informazione;
- con ogni domanda possiamo ottenere 1 bit di informazione.

Approfondimento

Sciuto ed altri, "Introduzione ai sistemi Informatici", McGraw Hill, 2005, terza edizione:

- gli argomenti trattati nel corso relativi alla codifica dell'informazione, ma in ordine inverso (capitolo 2, pagg. 17–55);
- cenni di teoria della trasmissione (capitolo 4, pagg. 113–133).