

# **Teoria dell'informazione e della trasmissione**

# Indice

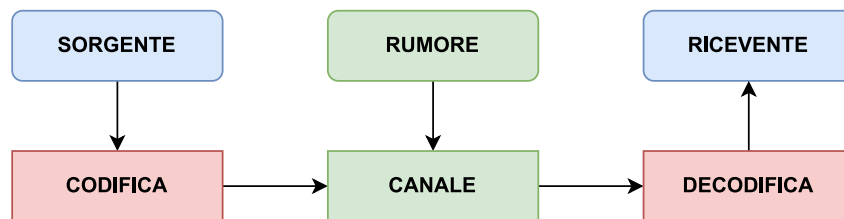
<b>1. Lezione 01 [24/09]</b> .....	<b>3</b>
<b>2. Lezione 02 [01/10]</b> .....	<b>4</b>
2.1. Introduzione storica .....	4
2.2. Cosa faremo .....	5
2.3. Esempio: informazione di un messaggio .....	5
2.4. Esempio: codice ZIP .....	5
2.5. Richiami matematici .....	6

## **1. Lezione 01 [24/09]**

## 2. Lezione 02 [01/10]

### 2.1. Introduzione storica

Lo schema di riferimento che andremo ad usare durante tutto il corso è il seguente:



La prima persona che lavorò alla teoria dell'informazione fu **Claude Shannon**, un impiegato della TNT (Telecom americana) al quale è stato commissionato un lavoro: massimizzare la quantità di dati che potevano essere trasmessi sul canale minimizzando il numero di errori che poteva accadere durante la trasmissione.

Nel 1948 pubblica un lavoro intitolato "A mathematical theory of communication", un risultato molto teorico nel quale modella in maniera astratta il canale e capisce come l'informazione può essere spedita "meglio" se rispetta certe caratteristiche. Ci troviamo quindi di fronte ad un risultato che non ci dice cosa fare nel caso specifico o che codice è meglio, ma è probabilistico, ti dice nel caso medio cosa succede.

Questo approccio è sicuramente ottimale, ma rappresenta un problema: va bene il caso medio, ma a me piacerebbe sapere cosa succede nel caso reale. Questo approccio più reale è quello invece seguito dal russo **Kolmogorov**, un accademico che vuole capire cosa succede nei singoli casi senza usare la probabilità. A metà degli anni '60 propone la sua idea di teoria dell'informazione, focalizzandosi quindi sui casi reali e non sui casi medi.

Questi due mostri sacri della teoria dell'informazione, nel nostro schema di lavoro, si posizionano nei rettangoli di sorgente e codifica, mentre la teoria della trasmissione si concentra sui rettangoli sottostanti.

Altri due personaggi che hanno lavorato allo stesso problema di Kolmogorov sono due russi, che lavoravano uno in America e uno nell'est del mondo, ma non sono ricordati perché Kolmogorov era il più famoso dei tre.

Un altro personaggio importante è **Richard Hamming**, un ricercatore di Bell Lab che doveva risolvere un problema: i job mandati in esecuzione dalle code batch delle macchine aziendali se si piantavano durante il weekend potevano far perdere un sacco di tempo. La domanda che si poneva Hamming era "che maroni, perché se le schede forate hanno errori, e le macchine lo sanno, io non posso essere in grado di correggerli?"

Lui sarà il primo che, per risolvere un problema pratico, costruisce il primo codice di rilevazione e correzione degli errori, il famoso e usatissimo **codice di Hamming**.

Vediamo alcune date che rivelano quanto è stata importante la teoria dell'informazione:

- 1965: prima foto di Marte in bianco e nero grande  $\approx 240K$  bit, inviata con velocità 6 bit al secondo, ci metteva ore per arrivare a destinazione;
- 1969: stessa foto ma compressa, inviata con velocità 16K bit al secondo, ci metteva pochi secondi;
- 1976: prima foto di Marte a colori da parte di Viking;
- 1979: prima foto di Giove e delle sue lune a colori da parte di Voyager;

- anni '80: prima foto di Saturno e delle sue lune da parte di Voyager.

## 2.2. Cosa faremo

Nel corso vedremo due operazioni fondamentali per spedire al meglio un messaggio sul canale:

- **compressione:** dobbiamo ottimizzare l'accesso al canale, ovvero se possiamo mandare meno bit ma questi mi danno le stesse informazioni dei bit totali ben venga;
- **aggiunta di ridondanza:** dobbiamo aggiungere dei bit per permettere il controllo, da parte del ricevente, dell'integrità del messaggio.

La compressione riguarda la **source coding**, e viene rappresentata nel **primo teorema di Shannon**, mentre la ridondanza riguarda la **channel coding**, e viene rappresentata nel **secondo teorema di Shannon**.

Voglio massimizzare l'informazione da spedire sul canale ma alla quale devo aggiungere ridondanza.

Quello che fa Shannon è modellare il canale secondo una matrice stocastica, che quindi permette di sapere in media cosa succede evitando tutti i casi precisi.

IN/OUT	$a$	$b$	$c$	$d$	$e$
$a$	0.7	0.0	0.1	0.1	0.1
$b$	0.2	0.8	0.0	0.0	0.0
$c$	0.1	0.0	0.6	0.2	0.1
$d$	0.0	0.0	0.2	0.5	0.3
$e$	0.0	0.0	0.0	0.0	1.0

Questa matrice indica, per ogni carattere del nostro alfabeto, quale è la probabilità di spedire tale carattere e ottenere i vari caratteri presenti nell'alfabeto.

## 2.3. Esempio: informazione di un messaggio

Quando un messaggio contiene più informazioni di un altro?

Lancio  $n$  volte due monete, una truccata e una non truccata.

Quale delle due monete mi dà più informazioni? Sicuramente quella non truccata: il lancio della moneta "classica" è un evento randomico, non so mai cosa aspettarmi, mentre il lancio della moneta truccata è prevedibile, so già cosa succederà.

Possiamo quindi dire che un evento prevedibile porta pochissima informazione, mentre un evento imprevedibile porta tantissima informazione.

## 2.4. Esempio: codice ZIP

Uno dei codici di compressione più famosi è il codice **ZIP**.

Come funziona:

1. creo un dizionario, inizialmente vuoto, che contiene le coppie (stringa-sorgente, codifica), dove "codifica" è un numero;
2. usando un indice che scorre la stringa carattere per carattere, e partendo con numero di codifica uguale a 1, fino all'ultimo carattere eseguo iterativamente:
  1. parti da una stringa vuota che useremo come accumulatore;

2. aggiungi il carattere corrente all'accumulatore;
3. se l'accumulatore non è presente come chiave nel dizionario la aggiungo a quest'ultimo con codifica il numero corrente di codifica; riparto poi dal punto 2.1 con numero di codifica aumentato di 1;
4. se l'accumulatore è presente come chiave nel dizionario riparto dal punto 2.2.

Ad esempio, la stringa *AABACDABDCAABBA* viene codificata con:

- $A \rightarrow 1$ ;
- $A \rightarrow AB \rightarrow 2$ ;
- $A \rightarrow AC \rightarrow 3$ ;
- $D \rightarrow 4$ ;
- $A \rightarrow AB \rightarrow ABD \rightarrow 5$ ;
- $C \rightarrow 6$ ;
- $A \rightarrow AA \rightarrow 7$ ;
- $B \rightarrow 8$ ;
- $B \rightarrow BA \rightarrow 9$ .

Prima avevo 15 caratteri, ora ne uso 9, letsgosky.

## 2.5. Richiami matematici

In questa parte abbiamo rivisto la definizione di monoide, gruppo, anello e campo, oltre alla definizione di generatore.

Abbiamo anche visto i **campi di Galois**, utilissimi per rendere il nostro calcolatore un campo algebrico, ovvero il campo  $\mathbb{GF}(2^{64})$  dei polinomi di grado massimo 63 con coefficienti sul campo  $\mathbb{Z}_2$ .