Teoria dell'informazione
Lezione I

Andrea Cosentino

31 December 2023

1 Introduzione

Durante il corso del '900 ci sono state diverse persone che hanno sviluppato le fondamenta della disciplina. Tra queste ricordiamo quelli che vengono considerati i padri della disciplina:

- Claude Shannon(USA): primo in assoluto. Fa una definizione in media
- Kolmogorov(URSS): arriva dopo Shannon ma fa una definizione puntuale. Espande il suo lavoro.
- Chaitin e Solomonoff: arrivano allo stesso tempo di Kolmogorov ma non vengonono considerati perché Kolmogorov era, ed è, più importante a livello accademico.

Durante questo corso ci proponiamo di riuscire a spedire dei dati da una sorgente a una destinazione attraverso un canale che può essere affetto da rumore.

Obiettivi del corso:

- Sfruttare al massimo il canale
- Gestire i bit persi nella trasmissione

2 Una visione d'insieme

(Sarebbe carino mettere un disegnino qui)
Shannon modella l'ambiente come composto da 3 attori:

- **Sorgente**: La sorgente genera il messaggio, lo codifica e lo spedisce sul canale.

- **Canale**: Il canale è il tramite tra la sorgente e la destinazione. E' il "posto" in cui passa l'informazione. E' affetto da **rumore**.
- **Ricevente**: Riceve il messaggio codificato. E' suo compito riuscirlo a decodificare.

Vogliamo codificare messaggi sorgente

(A) Massimizzando informazioni trasmesse A OGNI utilizzo del canale (problema di **Source coding**)

(B) Minimizzando, simultaneamente al primo punto, il numero di errori di trasmissione dovuti al rumore (problema di **Channel coding**). Shannon cerca di risolvere questo problema usando l'approccio divide et impera. Approccio che non è detto sia quello giusto. Infatti la soluzione ottimale dei due sottoproblemi non è detto che, se messe assieme, diano la soluzione ottimale per il problema. Questo perché potremmo non sfruttare possibili vantaggi di un problema sull'altro.
Questo non è il caso e infatti vale il seguente teorema.

TEOREMA di codifica sorgente e canale

L'unione delle soluzioni di source coding e channel coding (quindi dei due sottoproblemi risolti come indipendenti) dà la soluzione ottima.

Come risolvo il source coding? Enunciamo, in maniera non formale per adesso, il primo teorema di Shannon.

TEOREMA I teorema di Shannon

Si può comprimere, tramite un codice, un messaggio con perdite di informazioni piccole

Questo è dovuto al fatto che l'informazione non è uniformemente distribuita. Ci sono parti della codifica inutile.

Esempio:
Se codifico un cielo tutto azzurro, non ho bisogno di dire che ogni pixel è di colore azzurro, ma mi basta dire che una porzione di una foto è tutta azzurra.
La codidfica, cioè la rimozione della ridondanza, va ad amplificare il problema del rumore. Ogni bit perso è significativo. Per risolvere il problema di channel

coding useremo il secondo teorema di Shannon. Anche questo lo riportiamo, per ora, in modo informale.

TEOREMA II teorema di Shannon

Posso trasmettere con possibilità di errore piccola a piacere. Uitlizzo una ridondanza, controllata in base alla distorsione del canale.

3 Modellazione

Cominciamo a dare una definizione formale dei vari strumenti che utilizzeremo durante il corso.
Innanzitutto vederemo il canale come una matrice stocastica, cioè una matrice tale che la somma dei contributi di una riga è pari 1.
Per esempio, data questa matrice che rappresenta un canale

IN/OUT	a	b	c	d	e
a	0.7	0	0.1	0.1	0.1
b	0	0.5	0.5	0	0
c	0.1	0.1	0.1	0.1	0.6
d	0.2	0.1	0.3	0.1	0.3
e	0.4	0.2	0.2	0.1	0.1

Sulla prima riga sono presenti tutti i simboli che la destinazione può ricevere (a,b,c,d ed e), mentre sulla prima colonna tutti i simboli che può generare la sorgente. Il numero che si trova nella posizione (i,j) indica la probabilità che la sorgente generi, e invii, il simbolo i-esimo e che il ricevente ricevi il simbolo j-esimo.
Nel nostro caso, la probabilità che inviando a si riceva c è di 0.1, cioè 10%
Si noti come la matrice identità modelli un canale "perfetto" ovvero senza distorsione (rumore).

Nella matrice appaiono i simboli "a,b,c,d,e". Questi sono i simboli prodotti dalla sorgente. I simboli prodotti dalla sorgente appartengono a $\mathbb{X}$.
I messaggi sono definiti come segue:

Sia $\mathbb{X}$ l'insieme finito di simboli che compongono i messaggi generati dalla sorgente.

Un messaggio $x = (x_1, ..., x_n)\, \epsilon\, \mathbb{X}^n$ di lunghezza n è una sequenza di n simboli sorgente.

I simboli sorgente sono poi tradotti (quindi codificati) in parole di codice prima di essere inviati sul canale.

Una **parola di codice** è una sequenza di numeri dall'insieme 0,...,d-1 dei simboli di codice, dove d ¿ 1 è la base del codice.
Per effettuare la traduzione viene usata una **funzione di codifica**, che mappa i simboli sorgente in parole di codice

$$c : \mathbb{X} \to \{0, ..., d-1\}^+$$

Dove $\{0, ..., d-1\}^+$ è formalmente

$$\bigcup_{n=1}^{+\infty} \{0, ..., d-1\}^n$$

L'obiettivo che ci poniamo è di **minimizzare** $l_c(x)$, ovvero la lunghezza della parola di codice per il simbolo $x \epsilon \mathbb{X}$.

Risulta naturale cercare di assegnare a dei simboli che sono usati più spesso una parola di codice con lunghezza minore. Viceversa, a simboli usati raramente associamo lunghezze maggiori. Questo perché il nostro obiettivo è di minimizzare la lunghezza media pesata per la probabilità di utilizzo del simbolo, in poche parole il valore atteso.

Shannon definisce come $p(x)$ la probabilità di generazione di un simbolo. Inoltre assume, per semplicità, l'indipendenza di un simbolo dall'altro. Ovvero, la generezione di un simbolo $x \epsilon \mathbb{X}$ non influenza la generazione successiva di un simbolo $y \epsilon \mathbb{X}$

Definiamo la variabile casuale $X : \mathbb{X} \to \mathbb{R}$. Questa rappresenta l'estrazione di 1 simbolo dalla sorgente.
p diventa quindi la distribuzione di probabilità dei simboli della sorgente, mentre definiamo P_n come $P_n(x_1, ..., x_n) = p(x_1)...p(x_n)$. Questo vale perché l'estrazioni sono indipendenti.
P_n è la distribuzione sui messaggi $\mathbb{X}^n$.

Per avere una notazione più compatta, definiamo $\mathbb{D}$ come l'insieme $\{0, .., d-1\}$ dei simboli di codice con base d. Quindi c può essere definita come

$$c : \mathbb{X} \to \mathbb{D}^+$$

4 Problema codifica sorgente

Visti gli strumenti precedenti, possiamo definire in modo formale il problema della codifica sorgente che a inizio lezione avevamo descritto in modo non rigoroso.

PROBLEMA

Dato un modello di sorgente $< \mathbb{X}, p >$ e una base d > 1, trovare un codice $c : \mathbb{X} \rightarrow \mathbb{D}^+$ tale che il valore atteso

$$\mathbb{E}[l_c] = \sum_{x \epsilon \mathbb{X}} l_c(x) p(x)$$

della lunghezza di parola di codice sia minimo.

Il problema, così formulato, si presta a una soluzione banale e inutile! Infatti, basta dire che $c(x) = 0$ per gni $x \, \epsilon \, \mathbb{X}$. Bisogna introdurre delle limitazioni.

5 Limitazioni

La prima limitazione che introduciamo è che **il codice deve essere non singolare**. Un codice $c : \mathbb{X} \rightarrow \mathbb{D}^+$ è non signolare se a simboli della sorgente corrispondono parole di codice distinte (funzione iniettiva).

Formalmente,

$$\forall x, x^{'} \, \epsilon \mathbb{X} : x \neq x^{'} \, vale \, c(x) \neq c(x^{'})$$