Teoria dei codici

Prof. Marco Moretti

Corso di Laura Ing. Informatica - Università di Pisa

Introduzione

Applicazioni dei codici nel mondo che ci circonda

- Messaggi possono essere codificati per vari motivi
  - *Compressione dell'informazione*: Comprimere l'informazione eliminando tutta la ridondanza e risparmiare banda o spazio di memoria;
  - *Crittografia*: Nascondere il contenuto di un messaggio ad utenti diversi da quello desiderato;
  - *Rivelazione o correzione di errore*: Viene aggiunta ridondanza ad hoc per aumentare la resistenza a rumore e ad interferenza.
- Primi esempi di codifica, già nel 1800:
  - Messaggi via telegrafo o via radio che non dovevano essere intercettati (comunicazioni via telegrafo ottico in Francia);
  - Codifica a protezione di errore non sui bit ma sulle parole nelle trasmissioni via cavo sottomarino: non possono essere trasmesse parole che differiscono tra loro di meno di due lettere.

Applicazioni dei codici nel mondo che ci circonda

- Codici per applicazioni commerciali
  - Codici a rivelazione di errore: ISBN, carte di credito, TCP (16 bit checksum), codice ASCII (1 bit checksum).
  - Codici a correzione di errore: Hard disk (RS), cd (RS), comunicazioni cellulari, comunicazioni satellitari.

Applicazioni dei codici nel mondo che ci circonda: ridondanza nella lingua italiana

Sneocdo uno sdtiuo dlel'Untisverà di Cabmbrige, non irmptoa cmoe snoo sctrite le plaroe, tutte le letetre posnsoo esesre al pstoo sbgalaito, è ipmtortane sloo che la prmia e l'umiltia letrtea saino al ptoso gtsiuo, il rteso non ctona, il cerlvelo è comquune semrpe in gdrao di decraifre tttuo qtueso coas, pcheré non lgege ongi silngoa ltetrea, ma lgege la palroa nel suo insmiee...

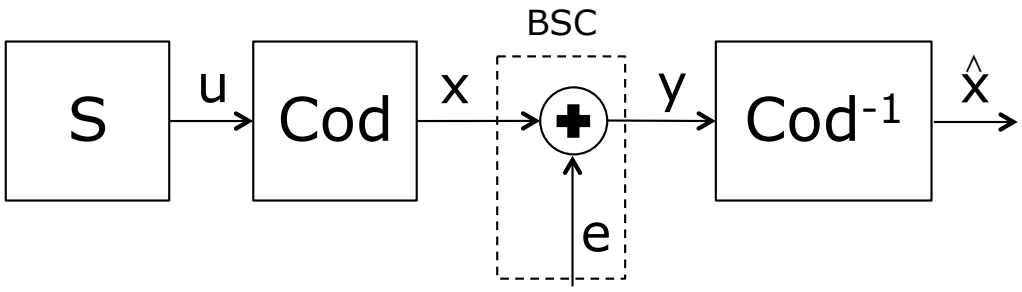
Storia della teoria dei codici

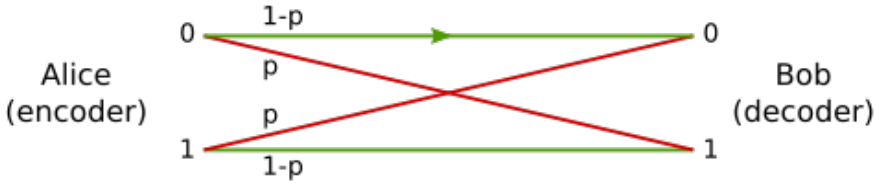
- Nel 1948 Claude Shannon scrive: "A Mathematical Theory of Communication" e nel 1949 "Communication Theory of Secrecy Systems".
  - Utilizza *bit* (binary digit) per descrivere la capacità di canale. La capacità C indica la massima quantità di informazione che può essere trasmessa in maniera affidabile su un dato canale.
  - La capacità dipende dalla banda B del canale e dal rapporto segnale rumore (signal-to-noise ratio, SNR)

$$C = B \log(1 + SNR)$$

Introduzione alla teoria dei codici

- Canale Gaussiano può essere modellato come binary symmetric channel (BSC) con probabilità di errore p.
  - Si assume che gli errori siano tra loro indipendenti.





Brevissima storia della teoria dei codici

- 1948 Articolo di Shannon;
- 1950 Codici di Hamming (codici a blocco);
- 1960 Codici Reed-Solomon (codici a blocco ciclici non binari);
- 1960 Codici LDPC;
- 1967 Algoritmo di Viterbi per la decodifica dei codici convoluzionali;
- 1968 Lancio sonda Mariner (decodifica soft);
- 1968 Lancio sonda Pioneer (codici convoluzionali);
- 1977 Lancio sonda Voyager (codici RS);
- 1993 Turbo codici;
- 1996 Codici convoluzionali per GSM;
- 2000 Applicazioni commerciali turbo codici;
- 2003 Applicazioni commerciali LDPC.

Tassonomia dei codici

- Codici lineari:
  - Codici a blocco
  - Codici convoluzionali
- Definizione di un codice a blocco
  - Rate $R = k/n$ del codice
    - Esempio: quale è il rate di un codice che ha 1024 parole di lunghezza 15 bit?
- Rivelazione di errore
- Correzione di errore

Un esempio di codici a blocco: codici a ripetizione

- I codici a ripetizione sono i codici a blocco più semplici che esistano.
- Di solito i codici a ripetizione codificano un solo bit alla volta e la codifica consiste nel ripetere il bit di ingresso $n-1$ volte.
- Ad esempio il codice a ripetizione con $R = 1/3$ ha solo due parole di codice:
  - $u = 0 \rightarrow x = [000]$
  - $u = 1 \rightarrow x = [111]$
- Il ricevitore effettua una decodifica a maggioranza: decide per il bit che compare nella maggioranza delle posizioni della parola ricevuta.
- Ad esempio:
  - $y = [000] \rightarrow \hat{x} = [000], \hat{u} = 0$
  - $y = [010] \rightarrow \hat{x} = [000], \hat{u} = 0$
  - $y = [101] \rightarrow \hat{x} = [111], \hat{u} = 1$

Un esempio di codici a blocco: codici a ripetizione

- Per una trasmissione sul BSC, la probabilità $p(t, n)$ di sbagliare t bit in una parola composta da n bit è

$$p(t, n) = \binom{n}{t} p^t (1-p)^{(n-t)},$$

dove il coefficiente binomiale

$$\binom{n}{t} = \frac{n!}{t!(n-t)!}$$

indica tutti i possibili pattern di errore cioè il numero di tutte le possibili combinazioni di t errori su n bit.

- Un codice a ripetizione con $R = 1/n$ può rivelare fino a $n-1$ errori e correggere $(n-1)/2$ errori (per n dispari).

Un esempio di codici a blocco: codici a ripetizione

- L'evento errore per un codice a correzione di errore consiste nel non essere in grado di correggere tutti gli errori introdotti dal canale.
- Se la probabilità di errore sul bit $p_{e,b}$ è sufficientemente piccola, la probabilità di errore $p_{e,W}$ per il codice può essere approssimata dal primo evento che determina la ricezione errata (ad esempio dall'aver fatto 2 errori per $R = 1/3$).
  - Codice a ripetizione $R = 1/3$
    1. Se $p_{e,b} = 0.1 \implies p_{e,W} \approx 2.7 * 10^{-2}$ $(p_{e,W} = 2.8 * 10^{-2})$
    2. Se $p_{e,b} = 0.01 \implies p_{e,W} \approx 2.97 * 10^{-4}$ $(p_{e,W} = 2.98 * 10^{-4})$.
  - Codice a ripetizione $R = 1/5$
    1. Se $p_{e,b} = 0.1 \implies p_{e,W} \approx 8.1 * 10^{-3}$
    2. Se $p_{e,b} = 0.01 \implies p_{e,W} \approx 9.8 * 10^{-6}$.

Un esempio di codici a blocco: codici a controllo di parità

- Codice con rate $R = k/(k+1)$: k bit informativi + 1 di parità (1 se #bit dispari, 0 altrimenti)

k	n	Stringa di bit	Bit di parità	Parola codificata
2	3	[10]	1	[101]
7	8	[1010101]	0	[10101010]

1. Esempio 1: Originariamente il codice ASCII era di 128 caratteri rappresentati da 7 bit, l'ottavo bit era un bit di controllo di parità così da essere tutto contenuto in un byte (codice ASCII esteso).
2. Esempio 2: Trasmetto parole di 11 bit con rate $R_b = 10Mb/s$ e probabilità di errore sul bit trasmesso $p_{e,b} = 10^{-8}$.
   - Senza controllo di parità è sufficiente che sia sbagliato anche un solo bit per sbagliare tutta la parola:

$$p_{e,W} = \sum_{j=1}^{11} \binom{11}{j} p_{e,b}^{j} (1-p_{e,b})^{(11-j)} \approx 11 p_{e,b} (1-p_{e,b})^{10} \approx 11p,$$

   ed il rate di parole sbagliate al secondo è

$$R_{e,w} = R_b/11 * p_{e,W} \approx (10^7/11) * 11p = 0.1w/s.$$

Un esempio di codici a blocco: codici a controllo di parità

- Esempio 2 -continuazione-
  - Aggiungo un bit di parità. La parola diventa di 12 bit e sbaglio quando faccio almeno 2 errori, gli errori di 1 bit vengono rivelati e corretti mediante la richiesta di ritrasmissione della parola.
    In questo caso, si ha

$$p_{e,W} = \sum_{j=2}^{12} \binom{12}{j} p_{e,b}^{j} (1-p_{e,b})^{(12-j)} \approx 66 p_{e,b}^{2} (1-p_{e,b})^{10} \approx 66 p_{e,b}^{2},$$

    ed il rate di parole sbagliate al secondo è

$$R_{e,w} = R_b/12 * p_{e,W} \approx (10^7/12) * 66 p_{e,b}^{2} = 5.5 * 10^{-9} w/s.$$

    Sbaglio una parola ogni $T_{e,w} = 1/R_{e,w} = 1.82 * 10^8$ s, una parola ogni sei anni circa (1 anno $\approx 3.15 * 10^7$s)!

Un esempio di codici a blocco: codice ISBN

- Il codice International Standard Book Number (ISBN) è un codice a controllo di parità per un alfabeto di simboli non binari. Ad ogni libro è assegnata una parola di codice di lunghezza $n = 10$ cifre in base decimale.
- Le prime nove cifre con indice $i = 1, 2, \ldots, 9$ identificano il libro, la decima con indice $i = 10$ è quella di controllo di parità e si calcola così:
  1. Si calcola la grandezza $z = \mathrm{mod}\,(S, 11)$ con

$$S = \sum_{j=1}^{9} (11 - j)x(j).$$

  2. La cifra di controllo di parità è il complemento a 11 di z

$$x(10) = \mathrm{mod}(11 - z, 11).$$

Solo per il controllo di parità, se $x(10) = 10 \implies x(10) = X$.

Un esempio di codici a blocco: codice ISBN

- Quando un dispositivo legge il codice ISBN, acquisisce la parola $y = [y(1), y(2), ..., y(10)]$.
  1. Per verificare che il codice sia corretto, il dispositivo calcola $\text{mod}\,(S', 11)$ con $S' = \sum_{j=1}^{10}(11-j)y(j)$,
  2. Assumendo che non ci siano errori su $x(10)$, si ha

$$\text{mod}\,(S', 11) = \text{mod}\left(\sum_{j=1}^{9}(11-j)y(j) + \text{mod}(11-z, 11), 11\right)$$

$$= \text{mod}\left(\sum_{j=1}^{9}(11-j)y(j) + \left(11 - \sum_{j=1}^{9}(11-j)x(j)\right), 11\right)$$

$$= \text{mod}\left(\sum_{j=1}^{9}(11-j)(y(j)-x(j)), 11\right)$$

Un esempio di codici a blocco: codice ISBN

- Se non ci sono errori si ha $y = x$ e quindi $\mathrm{mod}(S', 11) = 0$.
- Il codice è in grado di rivelare tutti gli errori singoli
  Sia $e(k)$ l'errore in posizione k, $y(k) = x(k) + e(k)$

$$\begin{aligned}\mathrm{mod}\ (S', 11) =& \mathrm{mod}\,((y(k) - x(k))(11 - k), 11) \\ &+ \mathrm{mod}\,(e(k)(11 - k), 11) \neq 0\end{aligned}$$

- Il codice è in grado di rivelare tutti i casi in cui ci sia uno scambio di due cifre del codice. Siano k_1 e k_2 le 2 posizioni scambiate

$$\begin{aligned}&\mathrm{mod}\ (S', 11) \\ &= \mathrm{mod}\,((x(k_2) - x(k_1))(11 - k_1) + (x(k_1) - x(k_2))(11 - k_2), 11) \\ &= \mathrm{mod}\,((x(k_2) - x(k_1))(k_2 - k_1), 11) \neq 0\end{aligned}$$