Codici Lineari

Nel seguito, indicheremo con $\mathbb{F}$ un generico campo finito. Come noto, $\mathbb{F}$ potrebbe essere l'insieme delle cifre binarie $\mathbb{F}_2 = \{0, 1\}$ con le usuali operazioni di prodotto e somma modulo 2. Più in generale, $\mathbb{F}$ è un campo di Galois con q elementi [1]. Nella codifica a blocco, vettori di K simboli della sorgente, appartenenti a $\mathbb{F}$, vengono mappati in vettori di N simboli. In un codice lineare, i vettori di codice costituiscono un sottospazio $\mathbf{C}$ di dimensione K dello spazio vettoriale N-dimensionale $\mathbb{F}^N$. È possibile pertanto individuare una base di tale sottospazio, costituita da K vettori, e costruire una matrice $\mathbf{G}$ le cui K righe coincidono con i vettori della base e che viene detta *matrice generatrice* del codice. Dato un vettore riga di ingresso $\mathbf{u} = [u_1, ..., u_K]$, $u_i \in \mathbb{F}$, l'insieme delle parole di codice risulta $\mathbf{C} = \{\mathbf{c} = \mathbf{uG} : \mathbf{u} \in \mathbb{F}^K\}$.

Senza perdita di generalità, in sostanza effettuando una permutazione delle associazioni fra ingressi e parole di codice, si può supporre che la matrice $\mathbf{G}$ sia in forma sistematica, ovvero possa scriversi come

$$\mathbf{G} = [\mathbf{I}_K|\mathbf{P}], \tag{1}$$

dove $\mathbf{I}_K$ è la matrice identità di dimensione $K \times K$. Un codice la cui matrice generatrice sia nella forma standard (1), viene detto *codice sistematico*. L'utilità di una matrice generatrice sistematica risiede nel fatto che risulta molto conveniente ricostruire l'ingresso dalla parola di codice: i primi K simboli (simboli di informazione) coincidono infatti con l'ingresso, mentre i simboli rimanenti (simboli di ridondanza) rappresentano dei *controlli di parità.*

Il *peso di Hamming* $d_w(\mathbf{x})$ di un vettore $\mathbf{x} \in \mathbb{F}^N$ viene definito come il numero delle componenti diverse da zero del vettore. Si definisce *distanza di Hamming* $d_H(\mathbf{x}, \mathbf{y})$ di due vettori $\mathbf{x}$ e $\mathbf{y}$ come il numero delle componenti in cui tali vettori differiscono. In altre parole, la distanza di Hamming di due vettori coincide con il peso di Hamming della differenza[1] $\mathbf{x}-\mathbf{y}$. Come noto, le proprietà di correzione d'errore di un codice sono determinate dalla distanza di Hamming minima d_H fra due qualsiasi parole del codice. Si noti che, per la linearità del codice, $\mathbf{x} \in \mathbf{C}$ e $\mathbf{y} \in \mathbf{C}$ implica $\mathbf{x} - \mathbf{y} \in \mathbf{C}$, e quindi d_H coincide con il minimo peso di Hamming delle parole di codice non identicamente nulle. Assumendo errori indipendenti sui simboli che costituiscono una parola di codice, la decodifica a massima verosimiglianza, che minimizza la probabilità di errore per vettori di codice equiprobabili, associa ad un generico vettore ricevuto, non appartenente a $\mathbf{C}$ a causa di errori di trasmissione, il vettore di codice *più vicino*, ovvero a minima distanza di Hamming. Un codice con distanza minima d_H corregge quindi correttamente fino ad almeno $t = \lceil (d_H - 2)/2 \rceil$ errori.

Si definisce il *prodotto scalare* fra due vettori $\mathbf{x}$ e $\mathbf{y}$ in $\mathbb{F}^N$, indicato con $\langle \mathbf{x}, \mathbf{y} \rangle$, come

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^{N} x_i y_i, \quad \langle \mathbf{x}, \mathbf{y} \rangle \in \mathbb{F}.$$

Nella relazione precedente, somme e prodotti sono effettuate secondo le regole del campo finito $\mathbb{F}$. Due vettori che verificano $\langle \mathbf{x}, \mathbf{y} \rangle = 0$ si dicono *ortogonali*. È immediato verificare che il prodotto scalare così definito soddisfa le regole usuali del prodotto scalare definito sui vettori di $\mathbb{R}^N$, con l'importante differenza che $\langle \mathbf{x}, \mathbf{x} \rangle = 0$ non implica necessariamente $\mathbf{x} = 0$. Ad esempio, il vettore $\mathbf{x} = [0\,1\,1\,0]$ di $\mathbb{F}_2^4$ è un vettore non nullo che verifica $\langle \mathbf{x}, \mathbf{x} \rangle = 0$. In altre parole, tale vettore è ortogonale a sé stesso.

Si consideri dunque un codice lineare $\mathbf{C}$ corrispondente ad un sottospazio di $\mathbb{F}^N$ di dimensione K. Si definisce il codice duale $\mathbf{C}^\perp$ come l'insieme dei vettori di $\mathbb{F}^N$ ortogonali ai vettori di $\mathbf{C}$

$$\mathbf{C}^\perp = \{\mathbf{y} \in \mathbb{F}^N : \langle \mathbf{x}, \mathbf{y} \rangle = 0, \mathbf{x} \in \mathbf{C}\}.$$

[1] Si noti che, nel caso dei codici binari, in cui $\mathbb{F} = \mathbb{F}_2 = \{0, 1\}$, la differenza coincide con la somma binaria.

È immediato verificare che $\mathbf{C}^\perp$ è esso stesso un sottospazio vettoriale di $\mathbb{F}^N$. Inoltre, vale il seguente risultato.

Teorema 1 *Sia $\mathbf{C} \subset \mathbb{F}^N$ un codice lineare di dimensione K, e sia $\mathbf{G}$ la matrice generatrice del codice. Allora*

1. *$\mathbf{C}^\perp = Ker\,\mathbf{G}$, dove $\mathbf{G}$ è la trasformazione lineare $\mathbf{G} : \mathbb{F}^N \to \mathbb{F}^K$ definita dalla matrice $\mathbf{G}$ di dimensione $K \times N$. Con $Ker\,\mathbf{G}$ si indica il nucleo della trasformazione lineare $\mathbf{G}$.*
2. *$\mathbf{C}^\perp$ ha dimensione $N - K$ e $(\mathbf{C}^\perp)^\perp = \mathbf{C}$.*

Prova.

1. Mantenendo la convenzione di rappresentare gli elementi di $\mathbb{F}^N$ come vettori riga, $\mathbf{y} \in Ker\,\mathbf{G}$ se e solo se $\mathbf{G}\mathbf{y}^t = 0$. Si ha $\mathbf{y} \in \mathbf{C}^\perp$ se e solo se il vettore $\mathbf{y}$ è ortogonale ai vettori di $\mathbf{C}$. Dato che le righe di $\mathbf{G}$ sono una base di $\mathbf{C}$, si ha $\mathbf{y} \in \mathbf{C}^\perp$ se e solo se $\mathbf{G}\mathbf{y}^t = 0$.
2. Come noto dalla teoria elementare delle trasformazioni lineari fra spazi vettoriali di dimensione N e K, si ha $N = \dim(Im\,\mathbf{G}) + \dim(Ker\,\mathbf{G})$, dove $\dim(Im\,\mathbf{G})$ denota la dimensione dello spazio vettoriale immagine della trasformazione $\mathbf{G}$. Inoltre, il rango per colonne di $\mathbf{G}$ è uguale al rango per righe, e dunque $\dim(Im\,\mathbf{G}) = K$. Si ha ovviamente $\mathbf{C} \subset (\mathbf{C}^\perp)^\perp$, ma dovendo avere $\mathbf{C}$ e $(\mathbf{C}^\perp)^\perp$ la stessa dimensione K, si ha $(\mathbf{C}^\perp)^\perp = \mathbf{C}$.

■

Esempio 1. Si consideri il codice $\mathbf{C} = \{[0\,0\,0\,0], [1\,0\,1\,0], [0\,1\,0\,1], [1\,1\,1\,1]\}$ su $\mathbb{F}_2^4$. Il codice duale ha dimensione 2 e risulta evidentemente $\mathbf{C}^\perp = \mathbf{C}$. Una matrice generatrice del codice in forma sistematica è

$$\mathbf{G} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}.$$

Le parole di codice si ottengono dalle combinazioni lineari delle righe di $\mathbf{G}$, attraverso la relazione $\mathbf{x} = \mathbf{u}\mathbf{G}$, con $\mathbf{u} \in \mathbb{F}_2^2$.

Dato un codice $\mathbf{C}$, si consideri una matrice generatrice $\mathbf{H}$ di $\mathbf{C}^\perp$. Tale matrice ha $N - K$ righe che costituiscono una base per $\mathbf{C}^\perp$, e si ha evidentemente $\mathbf{G}\mathbf{H}^t = 0$. La matrice $\mathbf{H}$ viene detta *matrice di controllo di parità* del codice. Se la matrice $\mathbf{G}$ è nella forma sistematica (1), si verifica immediatamente che $\mathbf{H} = [-\mathbf{P}^t | \mathbf{I}_{N-K}]$ è una matrice di controllo di parità del codice, dove $-\mathbf{P}$ rappresenta una matrice i cui elementi solo gli opposti degli elementi di $\mathbf{P}$ rispetto all'addizione definita in $\mathbb{F}$. Ad esempio, in $\mathbb{F}_2$, si ha $-\mathbf{P} = \mathbf{P}$.

Un'osservazione interessante è che una matrice di controllo di parità $\mathbf{H}$ definisce una trasformazione lineare $S : \mathbb{F}^N \to \mathbb{F}^{N-K}$ il cui nucleo $Ker\,\mathbf{H}$, per quanto dimostrato nel Teorema 1, coincide con il codice $\mathbf{C}$. Possiamo pertanto caratterizzare il codice come l'insieme dei vettori $\mathbf{x}$ che verificano la relazione $\mathbf{s} = \mathbf{x}\mathbf{H}^t = 0$. In generale, per $\mathbf{y} \in \mathbb{F}^N$, il vettore $\mathbf{s} = \mathbf{y}\mathbf{H}^t$ viene detto *sindrome* di $\mathbf{y}$, e per quanto visto esso risulta identicamente nullo se e solo se $\mathbf{y} \in \mathbf{C}$.
Diamo la seguente definizione.
Definizione. Sia $\mathbf{C}$ un codice e $\mathbf{x}$ un qualsiasi vettore di $\mathbb{F}^N$. L'insieme

$$\mathbf{x} + \mathbf{C} = \{\mathbf{x} + \mathbf{y} : \mathbf{y} \in \mathbf{C}\} \tag{2}$$

definisce un *coset* di $\mathbf{C}$.

Un coset viene dunque ottenuto aggiungendo le $|\mathbb{F}|^K$ parole di codice ad un generico vettore $\mathbf{x}$ di $\mathbb{F}^N$, ed è pertanto un insieme con $|\mathbb{F}|^K$ elementi. I coset godono delle seguenti proprietà.

Teorema 2 *Sia* $\mathbf{C}$ *un codice di* $\mathbb{F}^N$ *di dimensione* K*.*

1. *Se* $\mathbf{y} \in \mathbf{x} + \mathbf{C}$*, allora i coset* $\mathbf{x} + \mathbf{C}$ *e* $\mathbf{y} + \mathbf{C}$ *coincidono.*
2. *Un generico vettore* $\mathbf{x} \in \mathbb{F}^N$ *appartiene ad un unico coset.*
3. *Ci sono esattamente* $|\mathbb{F}|^{N-K}$ *coset.*

Prova.

1. Se $\mathbf{y} \in \mathbf{x} + \mathbf{C}$, allora possiamo scrivere $\mathbf{y} = \mathbf{x} + \mathbf{c}$, dove $\mathbf{c} \in \mathbf{C}$ è una parola di codice. Data la linearità del codice, abbiamo $\mathbf{c} + \mathbf{C} = \mathbf{C}$ e dunque $\mathbf{x} + \mathbf{C} = \mathbf{y} + \mathbf{C}$.
2. Dato che il vettore nullo è un vettore di codice, allora il vettore $\mathbf{x}$ appartiene al coset $\mathbf{x} + C$. Supponiamo ora che $\mathbf{x}$ appartenga al coset $\mathbf{y} + C$. Si ha allora $\mathbf{x} = \mathbf{y} + \mathbf{c}$, $\mathbf{c} \in \mathbf{C}$ e dunque $\mathbf{y} = \mathbf{x} - \mathbf{c}$. Ne deriva che $\mathbf{y} \in \mathbf{x} + \mathbf{C}$ e dunque $\mathbf{y} + \mathbf{C} = \mathbf{x} + \mathbf{C}$.
3. Dato che ogni vettore $\mathbf{x} \in \mathbb{F}^N$ appartiene esattamente a un coset e un coset ha $|\mathbb{F}|^K$ elementi, ci sono esattamente $|\mathbb{F}|^N / |\mathbb{F}|^K = |\mathbb{F}|^{N-K}$ coset.

■

È possibile enumerare i coset costruendo una tabella con $|\mathbb{F}|^{N-K}$ righe: la prima riga della tabella contiene semplicemente i vettori del codice $\mathbf{C}$. La prima riga contiene dunque gli elementi del coset $\mathbf{e}_1 + \mathbf{C}$ associata al cosiddetto *coset leader* $\mathbf{e}_1 = 0$. Si sceglie poi come coset leader della riga successiva il vettore $\mathbf{e}_2$ di $\mathbb{F}^N$ di peso di Hamming minimo fra quelli non ancora inseriti nella tabella, ponendo nella seconda riga della matrice gli elementi del coset $\mathbf{e}_2 + \mathbf{C}$. Si prosegue analogamente per le righe successive, ciascuna associata ad un coset leader di peso di Hamming crescente. La tabella contiene per costruzione tutti gli elementi di $|\mathbb{F}|^N$, e ciascuno di essi compare, sulla base del Teorema 2, in una ed una sola soltanto delle righe della tabella. Possiamo pertanto stabilire una corrispondenza fra un generico vettore $\mathbf{x}$ e il coset leader $\mathbf{e}_i$ della riga della tabella che contiene $\mathbf{x}$.

Si ha inoltre il seguente risultato.

Teorema 3 *Sia* $\mathbf{C}$ *un codice di dimensione* K *e sia* $\mathbf{s} = \mathbf{x}\mathbf{H}^t$ *la sindrome di* $\mathbf{x} \in \mathbb{F}^N$*. Due vettori* $\mathbf{x}_1$ *e* $\mathbf{x}_2$ *con la stessa sindrome* $\mathbf{s} = \mathbf{x}_1\mathbf{H}^t = \mathbf{x}_2\mathbf{H}^t$ *appartengono al medesimo coset.*

Prova. Da $\mathbf{x}_1\mathbf{H}^t = \mathbf{x}_2\mathbf{H}^t$ ricaviamo $(\mathbf{x}_1 - \mathbf{x}_2)\mathbf{H}^t = 0$. Quindi $\mathbf{x}_1 - \mathbf{x}_2 = \mathbf{c}$, con $\mathbf{c} \in \mathbf{C}$. Pertanto, $\mathbf{x}_1 \in \mathbf{x}_2 + \mathbf{C}$.

■

L'insieme delle sindromi costituisce dunque un sottospazio vettoriale di dimensione $|\mathbb{F}|^{N-K}$, e gli elementi di tale sottospazio possono essere posti in corrispondenza biunivoca con i coset leader della tabella di enumerazione dei coset. Per inciso, tale tabella può essere usata per la correzione di errore associando il vettore ricevuto alla parola di codice più vicina. Infatti, se il vettore ricevuto $\mathbf{x}$ si trova nella riga della tabella corrispondente ad un coset leader $\mathbf{e}_i$, per cui possiamo scrivere $\mathbf{x} = \mathbf{e}_i + \mathbf{c}$, si verifica facilmente che il procedimento di costruzione della matrice garantisce che $\mathbf{c}$ sia la parola di codice più vicina a $\mathbf{x}$. Se esistesse infatti una diversa parola di codice $\hat{\mathbf{c}}$ più vicina a $\mathbf{x} = \mathbf{e}_i + \mathbf{c}$, potremmo scrivere

$$d_H(\mathbf{e}_i + \mathbf{c}, \mathbf{c}) > d_H(\mathbf{e}_i + \mathbf{c}, \hat{\mathbf{c}}),$$

ovvero

$$d_w(\mathbf{e}_i) > d_w(\mathbf{e}_i + \mathbf{c} - \hat{\mathbf{c}}).$$

Questo non è tuttavia possibile, in quanto $\mathbf{e}_i$ e $\mathbf{e}_i + \mathbf{c} - \hat{\mathbf{c}}$ appartengono allo stesso coset e, per costruzione, il coset leader viene scelto come il vettore di peso minimo fra quelli non ancora comparsi nella tabella. La regola di decodifica a minima distanza consiste dunque nel correggere $\mathbf{x}$ con il vettore di codice $\mathbf{c}$.

Si verifica facilmente che il procedimento indicato consente di correggere correttamente i soli vettori di errore corrispondenti ai coset leader (che sicuramente, come visto precedentemente, includono tutti i vettori di errore di peso minore o uguale a t). Sia infatti $\mathbf{e}$ un vettore di errore non corrispondente ad un coset leader, si supponga di trasmettere la parola di codice $\mathbf{c}$ e di ricevere $\mathbf{x} = \mathbf{e} + \mathbf{c}$. Come per ogni altro vettore, possiamo scrivere $\mathbf{e} = \mathbf{e}_i + \hat{\mathbf{c}}$, con $\mathbf{e}_i$ un opportuno coset leader. Avremmo dunque $\mathbf{x} = \mathbf{e}_i + \hat{\mathbf{c}} + \mathbf{c}$, per cui $\mathbf{x}$ verrebbe decodificato nella parola di codice $\hat{\mathbf{c}} + \mathbf{c} \neq \mathbf{c}$.

Esempio 2. Il codice di Hamming(7,4) è un codice lineare binario con matrice di generazione del codice

$$\mathbf{G} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}.$$

Il codice associa a un vettore di ingresso $\mathbf{u}$ di 4 bit una parola di codice $\mathbf{c}$ di 7 bit, secondo la tabella seguente.

u	**c**	**u**	**c**	**u**	**c**	**u**	**c**
0000	0000000	0100	0100110	1000	1000101	1100	1100011
0001	0001011	0101	0101101	1001	1001110	1101	1101000
0010	0010111	0110	0110001	1010	1010010	1110	1110100
0011	0011100	0111	0111010	1011	1011001	1111	1111111

È immediato verificare che la distanza minima del codice è 3, per cui il codice corregge correttamente 1 errore nel vettore ricevuto. La tebella di decodifica mediante la sindrome risulta:

s	**e**	**s**	**e**
101	1000000	100	0000100
110	0100000	010	0000010
111	0010000	001	0000001
011	0001000	000	0000000

Pertanto, se $\mathbf{x}$ è il vettore di 7 bit ricevuto, si calcola $\mathbf{s} = \mathbf{x}\mathbf{H}^t$, si seleziona nella tabella il vettore $\mathbf{e}$ corrispondente, e si ricostruisce $\hat{\mathbf{x}} = \mathbf{x} + \mathbf{e}$. Una semplice rete logica permette di effettuare le operazioni richieste.

Riferimenti bibliografici

[1] S. Lin and D. Costello, Jr, *Error Control Coding: Fundamentals and Applications,* Prentice-Hall, 1983.

[2] R. J. McEliece, D. J. C. MacKay, and J.-F. Cheng, "Turbo decoding as an instance of Pearl's 'belief propagation' algorithm," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 140–152, Feb. 1998.

[3] R. Gallager, *Low Density Parity Check Codes*, MIT Press, 1963.

[4] D. MacKay, "Good error-correcting codes based on very sparse matrices," *IEEE Trans. Inform. Theory,* vol. 45, pp. 399-431, March 1999.