(F1X) Teoria dell'Informazione e della Trasmissione

Codici istantanei e disuguaglianza di Kraft

Docente: *Nicolò Cesa-Bianchi* versione 8 marzo 2016

Riprendiamo il nostro obiettivo di questa prima fase: la ricerca di un codice sorgente ottimo. Ovvero, dato un modello di sorgente $\langle \mathcal{X}, p \rangle$ vogliamo trovare il codice $c : \mathcal{X} \to \mathcal{D}^+$ tale che il valore atteso

$$\mathbb{E}\big[\ell_c\big] = \sum_{x \in \mathcal{X}} \ell_c(x)\, p(x) \tag{1}$$

della lunghezza di parola di codice sia minimo.

Come abbiamo già osservato in precedenza, perché un qualunque codice c sia utilizzabile in pratica deve essere possibile il processo di decodifica di messaggi. Ovvero, data una parola di codice $\boldsymbol{y} \in \mathcal{D}^+$ deve essere possibile risalire al messaggio $\boldsymbol{x} \in \mathcal{X}^+$ tale che $C(\boldsymbol{x}) = \boldsymbol{y}$. Per evitare di avere un codice inutilizzabile come soluzione ottima, ci siamo quindi ristretti ai codici univocamente decodificabili. Ovvero, i codici la cui estensione non è singolare. Purtoppo, anche i codici univocamente decodificabili possono presentare dei problemi dal punto di vista dell'utilizzo pratico, come si vede dall'esempio seguente.

Esempio 1 *Dato* $\mathcal{X} = \{\heartsuit, \diamondsuit, \clubsuit, \spadesuit\}$*, si consideri il seguente codice binario univocamente decodificabile*

$$c(\heartsuit) = 10 \quad c(\diamondsuit) = 00 \quad c(\clubsuit) = 11 \quad c(\spadesuit) = 110\ .$$

Supponiamo ora di aver utilizzato l'estensione C *per codificare un messaggio* $\boldsymbol{x} \in \mathcal{X}^+$ *e di aver ottenuto la parola di codice* $110\ldots0$*. Durante il processo di decodifica, per capire se il primo simbolo del messaggio sorgente è* $\clubsuit$ *piuttosto che* $\spadesuit$ *dobbiamo verificare se il numero di zeri che segue* 11 *sia pari o dispari. Infatti, se il numero di zeri è pari allora il messaggio dev'essere della forma* $\clubsuit\diamondsuit\ldots\diamondsuit$*. Invece, se il numero di zeri è dispari, allora il messaggio dev'essere della forma* $\spadesuit\diamondsuit\ldots\diamondsuit$*.*

L'esempio precedente mostra che un codice univocamente decodificabile può essere tale che per cominciare a decodificare il primo simbolo di un messaggio dobbiamo aspettare di leggere l'ultimo simbolo della sua codifica. Esaminando il codice, ci accorgiamo che il problema sta nel fatto che la parola di codice per $\clubsuit$ è un prefisso della parola di codice per $\spadesuit$. Se nessuna parola di codice fosse prefisso di un'altra, allora potremmo decodificare i simboli sorgente mano a mano che riceviamo i simboli di codice.

Si noti che possiamo rimediare a questo problema riservando un simbolo di codice per separare le codifiche dei simboli sorgente in una parola di codice. Per esempio, il codice binario dell'Esempio 1 diventerebbe il seguente codice ternario

$$c(\heartsuit) = 102 \quad c(\diamondsuit) = 002 \quad c(\clubsuit) = 112 \quad c(\spadesuit) = 1102\ .$$

Però è evidente che questa soluzione compromette la compattezza del codice.

Un codice $c : \mathcal{X} \to \mathcal{D}^+$ è detto **istantaneo** se nessuna parola di codice è prefisso di un'altra. Per esempio, il codice binario

$$c(\heartsuit) = 0 \quad c(\diamondsuit) = 10 \quad c(\clubsuit) = 110 \quad c(\spadesuit) = 111 \ .$$

è istantaneo.

Chiaramente, un codice istantaneo è non singolare. Mostriamo ora che i codici istantanei sono un sottoinsieme di quelli univocamente decodificabili.

Fatto 2 *Se c è istantaneo allora è anche univocamente decodificabile.*

Dimostrazione. Sia $c : \mathcal{X} \to \mathcal{D}^+$ un qualunque codice e sia C la sua estensione. Senza perdita di generalità, possiamo assumere che c sia non singolare. Infatti, se c fosse singolare allora non sarebbe istantaneo. Dimostriamo che se c non è univocamente decodificabile allora c non può essere istantaneo. Se c non è univocamente decodificabile esistono due messaggi $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}^+$ distinti tali che $C(\boldsymbol{x}) = C(\boldsymbol{x}')$. Ci sono solo due modi in cui $\boldsymbol{x}$ e $\boldsymbol{x}'$ possono essere distinti: (1) un messaggio è prefisso dell'altro; (2) c'è almeno una posizione in cui i due messaggi differiscono. Per analizzare il primo caso, assumiamo per esempio che $\boldsymbol{x}'$ sia un prefisso di $\boldsymbol{x}$. Ma allora, dato che $C(\boldsymbol{x}) = C(\boldsymbol{x}')$ i restanti simboli di $\boldsymbol{x}$ dovrebbero essere mappati da c nella parola di codice vuota, il che non è possibile per la nostra definizione di codice.[1] Rimane quindi da analizzare il secondo caso: c'è una posizione i in cui $\boldsymbol{x}$ e $\boldsymbol{x}'$ differiscono per la prima volta, ovvero $x_i \neq x'_i$ e $x_j = x'_j$ per ogni $j = 1, \ldots, i-1$. Quindi abbiamo $c(x_j) = c(x'_j)$ per $j = 1, \ldots, i-1$ e $c(x_i) \neq c(x'_i)$ dato che c è non singolare. Ma allora l'unico modo perché $C(\boldsymbol{x}) = C(\boldsymbol{x}')$ è che $c(x_i)$ sia prefisso di $c(x'_i)$ o viceversa, il che contraddice l'ipotesi che c sia istantaneo. □

Abbiamo così stabilito una gerarchia fra le funzioni della forma $c : \mathcal{X} \to \mathcal{D}^+$. Ovvero,

$$\text{codici istantanei} \ \subset \ \text{codici univ. decodificabili} \ \subset \ \text{codici non singolari}$$

dove le inclusioni sono strette. Infatti gli esempi precedenti hanno mostrato che esistono codici non singolari che non sono univocamente decodificabili e codici univocamente decodificabili che non sono istantanei.

I codici istantanei soddisfano un'importante proprietà strutturale che li rende riconoscibili anche soltanto in base alle sole lunghezze delle parole di codice.

Lemma 3 (Disuguaglianza di Kraft) *Dati $\mathcal{X} = \{x_1, \ldots, x_m\}$, $D > 1$ e m interi $\ell_1, \ldots, \ell_m > 0$, esiste un codice istantaneo $c : \mathcal{X} \to \mathcal{D}^+$ tale che $\ell_c(x_i) = \ell_i$ per $i = 1, \ldots, m$ se e solo se*

$$\sum_{i=1}^{m} D^{-\ell_i} \leq 1 \ .$$

Dimostrazione. Cominciamo a dimostrare che dato c istantaneo, le lunghezze delle sue parole di codice obbediscono alla disuguaglianza di Kraft. Sia $\ell_{\max}$ la lunghezza massima delle parole di

[1] Anche se estendessimo la definizione di codice ammettendo codici che mappano simboli sorgente nella parola vuota, tali codici non sarebbero istantanei, in quanto la parola vuota è prefisso di ogni altra parola di codice.

c,

$$\ell_{\max} = \max_{i=1,\ldots,m} \ell_c(x_i) \ .$$

Si consideri l'albero D-ario completo di profondità $\ell_{\max}$. Possiamo posizionare ogni parola di codice di c su un nodo dell'albero seguendo dalla radice il cammino corrispondente ai simboli della parola. Dato che il codice è istantaneo, nessuna parola apparterrà al sottoalbero avente come radice un'altra parola. Possiamo quindi partizionare le foglie dell'albero in sottoinsiemi disgiunti $A_1, \ldots, A_m$, dove A_i è il sottoinsieme di foglie associato alla parola $c(x_i)$ —si veda la Figura 1.

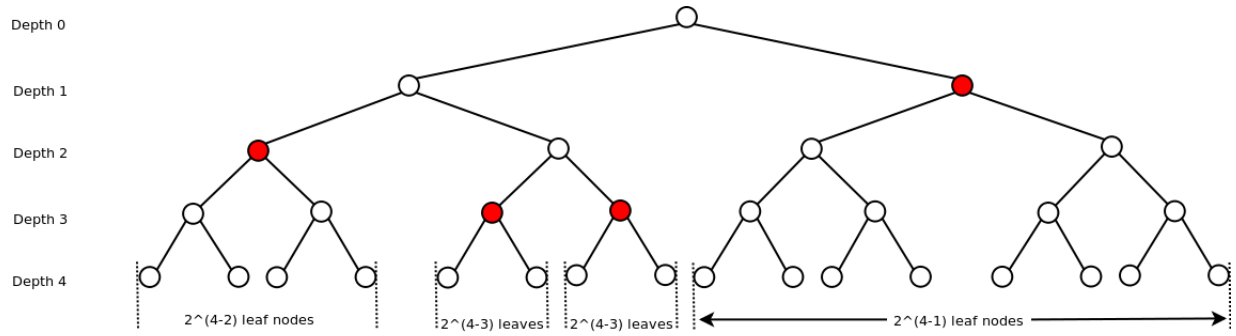

Figura 1: Partizione delle foglie indotta dal codice istantaneo indicato dai nodi colorati (da Wikipedia). La dimostrazione della disuguaglianza di Kraft può utilizzare un qualsiasi albero binario completo di profondità maggiore o uguale a $\ell_{\max}$. Possiamo quindi sostituire l'albero della figura con un albero binario completo di profondità $\ell_{\max} = 3$.

Ora, il numero di foglie nel sottoalbero di una parola ad altezza ℓ_i è ovviamente $D^{\ell_{\max}-\ell_i}$. D'altra parte, il numero totale di foglie nell'albero è $D^{\ell_{\max}}$. Quindi,

$$\sum_{i=1}^{m} D^{\ell_{\max}-\ell_i} = \sum_{i=1}^{m} |A_i| \leq D^{\ell_{\max}} \ .$$

Dividendo per $D^{\ell_{\max}}$ il membro sinistro e quello destro della formula otteniamo la disuguaglianza di Kraft.

Per dimostrare l'altra implicazione assumiamo che $\ell_1, \ldots, \ell_m > 0$ soddisfano la disuguaglianza di Kraft, e sia $\ell_{\max} = \max\{\ell_1, \ldots, \ell_m\}$. Allora possiamo costruire un codice istantaneo $c : \{x_1, \ldots, x_m\} \to \mathcal{D}^+$ con lunghezze date, ovvero $\ell_c(x_i) = \ell_i$ per $i = 1, \ldots, m$. A questo scopo si consideri l'albero D-ario ordinato e completo di profondità $\ell_{\max}$. Al simbolo x_1 associamo la parola di codice $c(x_1)$ corrispondente al nodo dell'albero di altezza ℓ_1 primo in ordine lessicografico (cioè più a sinistra). Ad ogni simbolo successivo x_i, associamo la parola di codice $c(x_i)$ corrispondente al primo nodo (sempre in ordine lessicografico) di altezza ℓ_i che né appartiene né include sottoalberi radicati su parole scelte in precedenza —si veda nuovamente la Figura 1. Si noti che il codice così costruito è istantaneo, dato che nessuna parola comparirà nel sottoalbero radicato su un'altra parola. Dato che le lunghezze soddisfano la disuguaglianza di Kraft, il numero totale di foglie necessarie a creare il codice è

$$\sum_{i=1}^{m} D^{\ell_{\max}-\ell_i} \leq D^{\ell_{\max}}$$

ovvero non maggiore delle foglie disponibili nell'albero. □

Vediamo ora come costruire un buon codice istantaneo, ovvero un codice che tenda a minimizzare (1). Fissato $D > 1$ e un modello di sorgente $\langle X, p \rangle$, la disuguaglianza di Kraft ci dice che possiamo limitarci a trovare dei numeri positivi $\ell_1, \ldots, \ell_m$ che la soddisfino. Infatti, trovati questi possiamo costruire automaticamente un codice istantaneo con quelle lunghezze. Quindi dobbiamo risolvere il problema

$$\begin{cases} \min\limits_{\ell_1,\ldots,\ell_m} \sum\limits_{i=1}^{m} \ell_i \, p_i \\ \text{tale che} \quad \sum\limits_{i=1}^{m} D^{-\ell_i} \leq 1 \end{cases}$$

dove $p_i = p(x_i)$ per $i = 1, \ldots, m$.

Possiamo ora osservare che, dato che $p_1 + \cdots + p_m = 1$ (è una distribuzione di probabilità), possiamo porre $D^{-\ell_i} \leq p_i$ in modo che

$$\sum_{i=1}^{m} D^{-\ell_i} \leq \sum_{i=1}^{m} p_i = 1$$

soddisfando così la disuguaglianza di Kraft. Risolvendo per ℓ_i otteniamo la condizione

$$\ell_i \geq \log_D \frac{1}{p_i} \, .$$

Quindi, è sufficiente porre $\ell_i = \left\lceil \log_D \frac{1}{p_i} \right\rceil$ per avere degli interi che soddisfano la disuguaglianza di Kraft. Il codice istantaneo risultante è noto come **codice di Shannon**.

È interessante studiare il caso particolare in cui i p_i siano dei reciproci di potenze di D. Per esempio, quando $p_1 = \frac{1}{2}$, $p_2 = \frac{1}{4}$, $p_3 = \frac{1}{8}$, $p_4 = \frac{1}{8}$ per $D = 2$ e $m = 4$. In questi casi, la lunghezza della codice di Shannon può essere calcolata esattamente come

$$\sum_{i=1}^{m} \ell_i \, p_i = \sum_{i=1}^{m} p_i \log_D \frac{1}{p_i} \, .$$

Questa quantità, che è una proprietà della distribuzione $p_1, \ldots, p_m$ soltanto, è nota come **entropia**. Le relazioni fra entropia e codifica ottima verrà approfondita nelle prossime lezioni.