Teoría de la Información y de la Transmisión

Códigos instantáneos y la desigualdad de Kraft

Docente: *Nicolò Cesa-Bianchi*
Traducción: *Mario Román*
 versión 28 de julio de 2016

Retomemos nuestro objetivo de esta primera fase: la búsqueda de un código fuente óptimo. Es decir, dado un modelo de fuente $\langle \mathcal{X}, p \rangle$ queremos encontrar el código $c : \mathcal{X} \to \mathcal{D}^+$ tal que el valor esperado

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

de la longitud de palabra de código sea mínimo.

como ya hemos observado antes, para que un código c cualquiera sea utilizable en la práctica, debe ser posible el proceso de decodificación de mensajes. Es decir, dada una palabra de código $\boldsymbol{y} \in \mathcal{D}^+$ debe ser posible remontarse al mensaje $\boldsymbol{x} \in \mathcal{X}^+$ tal que $C(\boldsymbol{x}) = \boldsymbol{y}$. Para evitar tener un código inutilizable como solución óptima, estamos por tanto restringidos a los códigos unívocamente decodificables. Es decir, los códigos cuya extensión no es singular. Por desgracia, también los códigos unívocamente decodificables pueden presentar problemas desde el punto de vista práctico, como se ve en el ejemplo siguiente.

Ejemplo 1 *Dado* $\mathcal{X} = \{\heartsuit, \diamondsuit, \clubsuit, \spadesuit\}$*, se considera el siguiente código binario unívocamente decodificable*

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

Supongamos ahora que hubiéramos utilizado la extensión C para codificar un mensaje $\boldsymbol{x} \in \mathcal{X}^+$ y hubiéramos obtenido la palabra de código $110\ldots0$*. Durante el proceso de decodificación, para entender si el primer símbolo del mensaje fuente es* ♣ *en vez de* ♠ *deberíamos verificar si el número de ceros que siguen a* 11 *es par o impar. De hecho, si el número de ceros es par, entonces el mensaje debe ser de la forma* $\clubsuit\diamondsuit\ldots\diamondsuit$*. En otro caso, si el número de ceros es impar, entonces el mensaje debe ser de la forma* $\spadesuit\diamondsuit\ldots\diamondsuit$*.*

El ejemplo precedente muestra que un código unívocamente decodificable puede ser tal que, para comenzar a decodificar el primer símbolo de un mensaje, debamos esperar a leer el último símbolo de su codificación. Examinando el código, nos damos cuenta de que el problema está en el hecho de que la palabra de código para ♣ es un prefijo de la palabra de código para ♠. Si ninguna palabra de código fuese prefijo de otra, entonces podríamos decodificar los símbolos fuente conforme recibiésemos los símbolos de código.

Nótese que podemos remediar este problema reservando un símbolo de código para separar las codificaciones de los símbolos fuente en una palabra de código. Por ejemplo, el código binario del Ejemplo 1 se convertiría en el siguiente código ternario

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

Sin embargo, es evidente que esta solución compromete la compacidad del código.

Un código $c : \mathcal{X} \to \mathcal{D}^+$ se llama **instantáneo** si ninguna palabra de código es prefijo de ninguna otra. Por ejemplo, el código binario

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

es instantáneo.

Claramente, un código instantáneo es no singular. Mostramos ahora que los códigos instantáneos son un subconjunto de aquellos unívocamente decodificables.

Proposición 2 *si c es instantáneo entonces es también unívocamente decodificable.*

Dimostrazione. Sea $c : \mathcal{X} \to \mathcal{D}^+$ un código cualquiera y sea C su extensión. Sin pérdida de generalidad, podemos asumir que c sea no singular. De hecho, si c fuese singular, entonces no sería instantáneo. Demostramos que si c no es unívocamente decodificable, entonces c no puede ser instantáneo. Si c no es unívocamente decodificable existen dos mensajes $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}^+$ distintos tales que $C(\boldsymbol{x}) = C(\boldsymbol{x}')$. Solo hay dos formas en las que $\boldsymbol{x}$ y $\boldsymbol{x}'$ pueden ser distintos: (1) un mensaje es prefijo del otro; (2) hay al menos una posición en la que los dos mensajes difieren. Para analizar el primer caso, asumimos por ejemplo que $\boldsymbol{x}'$ es un prefijo de $\boldsymbol{x}$. Pero entonces, dado que $C(\boldsymbol{x}) = C(\boldsymbol{x}')$, los restantes símbolos de $\boldsymbol{x}$ deberían ser mapados desde c en la palabra de código vacía, lo que no es posible por nuestra definición de código.[1] Queda entonces analizar el segundo caso: hay una posición i en la cual $\boldsymbol{x}$ y $\boldsymbol{x}'$ difieren por primera vez, es decir $x_i \neq x'_i$ y $x_j = x'_j$ para cada $j = 1, \ldots, i-1$. Entonces tenemos $c(x_j) = c(x'_j)$ para $j = 1, \ldots, i-1$ y $c(x_i) \neq c(x'_i)$ dado que c es no singular. Pero entonces la única forma de que $C(\boldsymbol{x}) = C(\boldsymbol{x}')$ es que $c(x_i)$ sea prefijo de $c(x'_i)$ o viceversa, lo que contradice la hipótesis de que c es instantáneo. □

Hemos establecido así una jerarquía entre las funciones de la forma $c : \mathcal{X} \to \mathcal{D}^+$. Es decir,

$$\text{códigos instantáneos} \ \subset \ \text{códigos univ. decodificables} \ \subset \ \text{códigos no singulares}$$

donde las inclusiones son estrictas. De hecho, los ejemplos precedentes han mostrado que existen códigos no singulares que no son unívocamente decodificables y códigos unívocamente decodificables que no son instantáneos.

Los códigos instantáneos satisfacen una importante propiedad estructural que los mantiene reconocibles también sólo en base a las longitudes de las palabras de código.

Lema 3 (Desigualdad de Kraft) *Dados $\mathcal{X} = \{x_1, \ldots, x_m\}$, $D > 1$ y m enteros $\ell_1, \ldots, \ell_m > 0$, existe un código instantáneo $c : \mathcal{X} \to \mathcal{D}^+$ tal que $\ell_c(x_i) = \ell_i$ para $i = 1, \ldots, m$ si y sólo si*

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

[1] Incluso si entendiésemos la definición de código admitiendo códigos que mapan símbolos fuente en la palabra vacía, tales códigos no serían instantáneos, en cuanto la palabra vacía sería prefijo de cualquier otra palabra de código.

Dimostrazione. Comenzamos a demostrar que dado c instantáneo, las longitudes de sus palabras de código obedecen la desigualdad de Kraft. Ses $\ell_{\max}$ la longitud máxima de las palabras de c,

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

Considérese el árbol D-ario completo de profundidad $\ell_{\max}$. Podemos posicionar cada palabra de código de c sobre un nodo del árbol siguiendo desde la raíz el camino correspondiente a los símbolos de la palabra. Dado que el código es instantáneo, ninguna palabra pertenecerá al subárbol teniendo como raíz otra palabra. Podemos entonces particionar las hojas del árbol en subconjuntos disjuntos $A_1, \ldots, A_m$, donde A_i es el subconjunto de hojas asociado a la palabra $c(x_i)$ —véase la Figura 1.

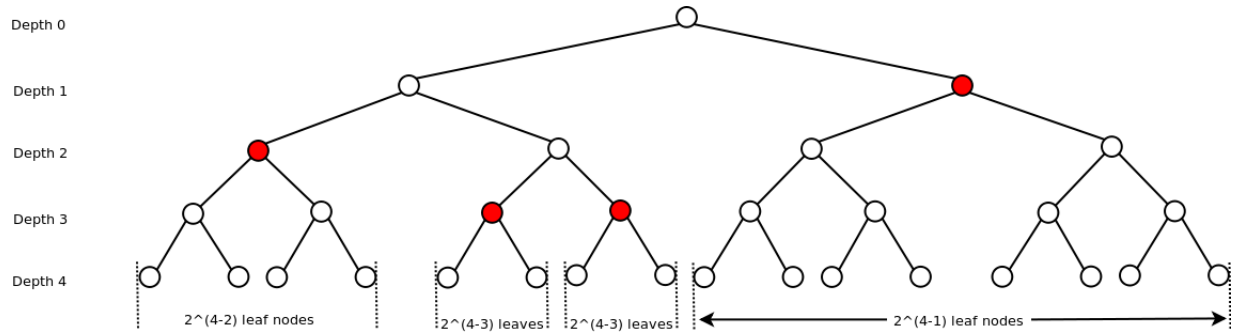

Figura 1: Partición de las hojas inducida desde el código instantáneo indicado por los nodos coloreados (de Wikipedia). La demostración de la desigualdad de Kraft puede utilizar un árbol binario completo cualquiera de profundidad mayor o igual a $\ell_{\max}$. Podemos entonces sustituir el árbol de la figura por un árbol binario completo de profundidad $\ell_{\max} = 3$.

Ahora, el número de hojas en el subárbol de una palabra de altura ℓ_i es obviamente $D^{\ell_{\max}-\ell_i}$. Por otro lado, el número total de hojas en el árbol es $D^{\ell_{\max}}$. Por tanto,

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

Dividiendo por $D^{\ell_{\max}}$ el miembro izquierdo y el de la derecha de la fórmula obtenemos la desigualdad de Kraft.

Para demostrar la otra implicación asumimos que $\ell_1, \ldots, \ell_m > 0$ satisfacen la desigualdad de Kraft, y sea $\ell_{\max} = \max\{\ell_1, \ldots, \ell_m\}$. Entonces podemos construir un código instantáneo $c : \{x_1, \ldots, x_m\} \to \mathcal{D}^+$ con longitudes dadas, es decir, $\ell_c(x_i) = \ell_i$ para $i = 1, \ldots, m$. A este propósito, considérese el árbol D-ario ordenado y completo de profundidad $\ell_{\max}$. Al símbolo x_1 le asociamos la palabra de código $c(x_1)$ correspondiente al nodo del árbol de altura ℓ_1 primero en orden lexicográfico (es decir, más a la izquierda). A cada símbolo sucesivo x_i, asociamos la palabra de código $c(x_i)$ correspondiente al primer nodo (siempre en orden lexicográfico) de altura ℓ_i que ni pertenezca ni incluya subárboles radicados sobre palabras elegidas previamente —véase nuevamente la Figura 1. Nótese que el código así construido es instantáneo, dado que ninguna palabra aparecerá en el subárbol radicado sobre otra palabra. Dado que las longitudes satisfacen la desigualdad de Kraft,

el número total de hojas necesarias para crear el código es

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

es decir, no mayor de las hojas disponibles en el árbol. □

Veamos ahora cómo construir un buen código instantáneo, es decir, un código que tienda a minimizar (1). Fijado $D > 1$ y un modelo de fuente $\langle X, p\rangle$, la desigualdad de Kraft nos dice que podemos limitarnos a buscar números positivos $\ell_1, \ldots, \ell_m$ que la satisfagan. De hecho, una vez encontrados, podemos construir automáticamente un código instantáneo con esas longitudes. Por tanto, debemos resolver el problema

$$\begin{cases} \min\limits_{\ell_1,\ldots,\ell_m} \displaystyle\sum_{i=1}^{m} \ell_i\, p_i \\ \text{tal que} \quad \displaystyle\sum_{i=1}^{m} D^{-\ell_i} \leq 1 \end{cases}$$

donde $p_i = p(x_i)$ para $i = 1, \ldots, m$.

Podemos ahora observar que, dado que $p_1 + \cdots + p_m = 1$ (es una distribución de probabilidad), podemos poner $D^{-\ell_i} \leq p_i$ de forma que

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

satisfaciendo así la desigualdad de Kraft. Resolviendo para ℓ_i obtenemos la condición

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

Por tanto, es suficiente tener $\ell_i = \left\lceil \log_D \frac{1}{p_i} \right\rceil$ para tener enteros que satisfagan la desigualdad de Kraft. El código instantáneo resultante se conoce como **código de Shannon**.

Es interesante estudiar el caso particular en el que los p_i sean inversos de potencias de D. Por Ejemplo, cuando $p_1 = \frac{1}{2}$, $p_2 = \frac{1}{4}$, $p_3 = \frac{1}{8}$, $p_4 = \frac{1}{8}$ para $D = 2$ e $m = 4$. En estos casos, la longitud del código de Shannon puede ser calculada exactamente como

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

Esta cantidad, que es una propiedad sólo de la distribución $p_1, \ldots, p_m$, se conoce como **entropía**. Las relaciones entre entropía y codificación óptima vendrán profundizadas en las siguientes lecciones.