Teoría de Información y la Comunicación
Clase 5

Temas de la Unidad:

- Inecuación de Kraft
- Inecuación de Mac Millan
- Longitud media de un código
- Códigos compactos

Sea ún alfabeto fuente $S = \{s_1, s_2, \ldots, s_q\}$ y un alfabeto código $X = \{x_1, x_2, .., x_r\}$. Consideremos un código $C = \{c_1, c_2, .., c_q\}$ donde cada c_i se expresa con una palabra código compuesto por una combinación de símbolos en X de longitud l_i.

Se dice que es condición **suficiente** para que exista al menos un código instantáneo con estas longitudes si se cumple que:

$$\sum_{i=1}^{i=q} r^{-l_i} \leq 1$$

Inecuación de Kraft

Donde r es el número de símbolos de X.

Si consideramos el alfabeto binario tendremos : $$\sum_{i=1}^{i=q} 2^{-l_i} \leq 1$$

La inecuación condiciona las longitudes de las palabras y no las palabras mismas. Sólo es una medida cuantitativa ya que dice que es posible pero no asegura que el código construido sea instantáneo. Luego, se debe verificar la condición de prefijo para asegurar que el código es instantáneo.

Unidad III: Propiedades de los Códigos

Dada una fuente S { s_1, s_2, s_3, s_4} y los codificaciones A,B,C,D,E :

Símbolos de S	A	B	C	D	E
S_1	00	0	0	0	0
s_2	01	100	10	100	10
s_3	10	110	110	110	110
s_4	11	111	111	11	11
Kraft	1	7/8 < 1	1	1	1 1/8 >1
	No hay Pref. Es Inst.	No hay Pref. Es Inst.	No hay Pref. Es Inst.	Hay Pref. No Inst.	No Inst.

Si se cumple la inecuación de Kraft, existe un código instantaneo con esas longitudes. **Pero no garantiza que cualquier código con esas longitudes sea instantaneo.** **Para verificar si es instantaneo es preciso que no existan prefijos en el código.**

Como los códigos unívocos incluyen a los instantáneos, Mac Millan demuestra que la inecuación de Kraft es condición **necesaria y suficiente** para la existencia de un código unívoco.

Esto significa que si se cumple la inecuación de Kraft-MacMillan:

$$\sum_{i=1}^{i=q} r^{-l_i} \leq 1$$

El código es unívoco con seguridad y que si luego se verifica la inexistencia de prefijos será instantáneo.

Ejemplo: Tenemos una fuente de S de 10 símbolos y queremos codificarla con un código instantáneo de tres símbolos o tri-nario (r=3). Verificar si es posible hacerlo con las longitudes: 1, 2, 2, 2, 2, 2, 3, 3, 3, 3.

Aplicando la inecuación de Kraft, se obtiene que: como es mayor que 1 no es posible definir un código instantáneo con esas longitudes.

$$\sum_{i=1}^{i=q} r^{-l_i} = \frac{28}{27} > 1$$

Ejemplo: Supongamos fuente S de nueve símbolos con un código instantáneo tri-nario (r=3), con palabras de longitudes: 1, 2, 2, 2, 2, 2, 3, 3, 3. Por Kraft-MacMillan tenemos: $\sum_{i=1}^{i=q} r^{-l_i} = 1$

En consecuencia el código es aceptable y se define como sigue: se elige un prefijo de longitud 1 (el 0), quedando obligados a adoptar uno de los dos restantes prefijos de longitud 1 para las demás palabras. Esto limita a dos veces tres, es decir seis, las palabras permitidas de longitud 2. Se emplean únicamente cinco de ellas, conservando la sexta (22) como prefijo de las tres últimas palabras.

S1	0	S6	21
S2	10	S7	220
S3	11	S8	221
S4	12	S9	222
S5	20	-----	

Dado un alfabeto fuente S y un alfabeto código X, definir un código consiste en establecer una regla por la cual a cada símbolo en S le corresponda una cadena, extensión o palabra código de X. Partiendo de esta definición infinidad de códigos son posibles, en nuestro caso nos interesan los unívocamente decodificables y en particular los instantáneos.
Un aspecto central en la codificación es establecer un criterio para poder seleccionar el "mejor" u "óptimo" en algún sentido.

Un código será óptimo por ser cómodo, fiable(frente a errores), seguro (frente a usuarios no autorizados), menos caro (almacenamiento y transmisión). Para evaluarlos se suele considerar más de un aspecto y en principio se utilizará sólo el criterio "cuanto más breve mejor". Esto es así porque la longitud del código afecta el almacenamiento y la velocidad de transmisión.

Para cuaquier código la longitud media es el promedio de las longitudes de sus palabras código ponderadas por sus probabilidades:

Sea un código bloque X con palabras código $\{X_1,X_2,\ldots,X_q\}$ con longitudes $\{l_1,l_2,..,l_q\}$.El cual representa al alfabeto fuente S $\{S_1,S_2,..,S_q\}$, con probabilidades P $\{p_1,p_2,..p_q\}$. Se llama longitud media del código (L) a: $L=\sum_{i=1}^{i=q} P_i * l_i$. La definición de L es válida tanto para las fuentes de memoria nula como de Markov.

Unidad III: Propiedades de los Códigos

Si asumimos que el mejor código es el mas breve, pues aprovecha el almacenamiento y la velocidad de transferencia. Entonces el mejor código será el más compacto, en otras palabras dado un codigo construido en un alfabeto código X, para un alfabeto S con r símbolos. Este será optimo o compacto si su longitud media es igual o menor que la longitud media de todos los códigos unívocos que pueden aplicarse a S utilizando X.

Ubicar un código unívoco decodíficable y compacto es un problema fundamental de la codificación y si este además debe ser instantáneo el problema es mayor.

Teniendo en cuenta la definición y propiedades de la Entropía y la inecuación de Kraft se puede demostrar que: $H_r(S) \leq L$. Lo que equivale a decir que con un código instantáneo y una fuente de memoria nula, L debe ser igual o mayor que $H_r(S)$ y alcanzará su mínimo valor si pueden elegirse las longitudes (l_i) de las palabras iguales a $\log_r(\frac{1}{p_i})$

Dado que se trata de longitudes de palabras, dicho logaritmo debe ser un número entero. Teniendo en cuenta la definición de logaritmo, las probabilidades P_i de los símbolos será de la forma: $(\frac{1}{r})^{\alpha_i}$. Donde α_i es un número entero. Si estas condiciones se cumplen se habrán encontrado las longitudes de las palabras que constituyen un código compacto.

Unidad III: Propiedades de los Códigos

Tenemos un alfabeto fuente S: $\{s_1,s_2,s_3,s_4\}$ con probabilidades P = {1/4,1/4,1/4,1/4}. Tenemos que $\sum_{i=1}^{i=4} p_i * \log_2(\frac{1}{(p_i)}) = 2\, binits/símbolo$, de esto se desprende que es imposible codificar los símbolos de esta fuente mediante un código binario unívoco de longitud media inferior de 2 binits por símbolo.

S_1	00
S_2	01
S_3	10
S_4	11

Como el alfabeto código tiene un r=2 y la probabilidad de cada símbolo s_i = 1/4, entonces $(\frac{1}{r})^{\alpha_i}$ Da como resultado $1/4=(1/2)^2$. Como α_i = 2 un código compacto deberá tener palabras de longitud 2.