Tarea 3. Compresión de datos.

Septiembre 27, 2004

1. Dado un número a en el intervalo $[0, 1)$ con representación $[b_1, b_2, ..., b_n]$, muestre que para tener una representación binaria de b con $[b_1, b_2, ..., b_n]$ como prefijo, b tiene que estar en el intervalo $[a, a + \frac{1}{2})$.

   Sea S_n definida como:
$$S_n = \sum_{i=0}^{n} r^i,$$
   donde $r \geq 0$. Si $r = 0$ entonces $S_n = 0 \ \forall n$. Si $r > 0$ entonces tenemos que:
$$rS_n = \sum_{i=1}^{n+1} r^i \Rightarrow rS_n + 1 = S_{n+1}$$
   pero
$$S_{n+1} - S_n = r^{n+1}$$
   por tanto:
$$rS_n + 1 = S_{n+1} \Rightarrow rS_n + 1 - S_n = S_{n+1} - S_n = r^n \Rightarrow S_n(r-1) = r^n - 1 \Rightarrow S_n = \frac{r^n - 1}{r - 1}.$$
   Si $0 < r < 1$ entonces
$$\lim_{n \to \infty} S_n = \frac{1}{1 - r}.$$
   Entonces se cumple que:
$$\sum_{i=n+1}^{\infty} r^i = \frac{1}{1-r} - \sum_{0}^{n} r^i = \frac{1}{1-r} - S_n = \frac{1}{1-r} - \frac{1 - r^{n+1}}{1-r} = \frac{r^{n+1}}{1-r}$$
   En particular para $r = \frac{1}{2}$:
$$\lim_{n \to \infty} S_n = \sum_{i=0}^{\infty} \left(\frac{1}{2}\right)^i = 2$$
   y
$$\sum_{i=n+1}^{\infty} \left(\frac{1}{2}\right)^i = \frac{\left(\frac{1}{2}\right)^{n+1}}{\frac{1}{2}} = \left(\frac{1}{2}\right)^n.$$
   Es claro que cualquier *suma parcial* S_n es menor que $\sum_{i=0}^{\infty} r^i$.

   Sea $0 \leq b < 1$ un número con representación binaria $[b_0, b_1, ..., b_m]$, con $n \geq m$, entonces $b_0 = 0$ y $b = \sum_{i=0}^{m} b_i r^i$, pero
$$a = \sum_{i=0}^{n} b_i r^i \Rightarrow b = a + \sum_{i=n+1}^{m} b_i r^i \leq a + \sum_{i=n+1}^{m} r^i < a + \sum_{i=n+1}^{\infty} = a + \left(\frac{1}{2}\right)^n.$$
   Por tanto $b \in \left[a, a + \left(\frac{1}{2}\right)^n\right)$

2. Dado el modelo de probabilidad con función de densidad $p(a_1) = 0.2$, $p(a_2) = 0.3$, $p(a_3) = 0.5$

   (a) Encuentre la etiqueta para la secuencia $a_1 a_1 a_3 a_2 a_3 a_1$
   La etiqueta para esta secuencia es la representación binaria de cualquier número en el intervalo $(0.027, 0.0276)$. *La mínima cantidad de bits requeridos es* 8. *La secuencia* 00000111 *representa el número* 0.02734375 $\in (0.027, 0.0276)$

   (b) Decodifique la secuencia con etiqueta 0.63215699
   El código correspondiente es $a_3 a_2 a_2 a_1 a_2 a_1$

3. Genere una secuencia binaria de longitud L con $P(0) = 0.8$ y use el algoritmo de código aritmético para decodificarla. Grafique la diferencia de la razón $\frac{bits}{simbolo}$ y la entropía como función de L. Comente sobre el efecto de L en la razón $\frac{bits}{simbolo}$.

   La gráfica y comentarios se muestran por separado en la página...

4. Implementar el algoritmo LZW. Mostrar la ganancia en compresión y el diccionario correspondiente.

   Los resultados se muestran por separado en la página...