

MA37A Optimización. Semestre Primavera 2006

Profesor: Héctor Ramírez C. Auxiliares: Rodrigo Lopez, Oscar Peredo.

Control 3

P1. (a) (Web Mining Estructural)

Una página web se considera *importante* si es apuntada, mediante *hyperlinks*, por otras páginas web. Una medida de importancia entre páginas web es el *PageRank*, el cual se calcula de la siguiente forma:

$$\text{PageRank}(p) = (1 - \alpha) + \alpha \cdot \sum_{q \text{ apunta a } p} \left(\frac{\text{PageRank}(q)}{c(q)} \right)$$

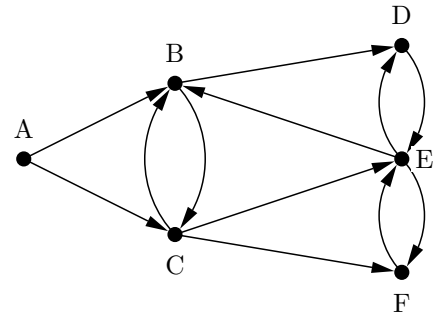
donde $\alpha \in [0, 1]$ es un factor de amortiguamiento y $c(q)$ es el número de links salientes de la página q .

Sea un grafo de páginas y links como se muestra en la figura. Se sabe que $\text{PageRank}(A) = 0$, $\text{PageRank}(B) = 2$ y $\text{PageRank}(E) = 5$. Si se considera $\alpha = 1$ y que el peso de un arco (p, q) es calculado como el promedio de los *PageRank* de p y q , i.e. $(\text{PageRank}(p) + \text{PageRank}(q))/2$:

(i) Muestre que $\text{PageRank}(C) = 1$, $\text{PageRank}(D) = 8/3$ y $\text{PageRank}(F) = 2$. Calcule el peso de los arcos del grafo.

(ii) Calcule el camino de peso mínimo entre las páginas A y E .

Indicación: Al calcular los pesos de los arcos, puede serle útil realizar una ponderación para que estos valores queden enteros.



(b) Resuelva el siguiente problema lineal entero usando el algoritmo de ramificación y acotamiento (RyA):

$$\text{máx } p = 3x + 2y + z ; x + y \leq 10, x + 3y - 2z \geq 15, 2x - y + z \leq 9, x \geq 0, y \in \mathbb{N} \cup \{0\}, z \in \{0, 1, 2\} \quad (\text{PLE})$$

Indicaciones:

1) La solución de la versión lineal *relajada* de (PLE) problema está dada por: $p = 55/2$; $x = 23/4$, $y = 17/4$, $z = 7/4$.

2) Para mejorar la eficiencia del algoritmo RyA es fuertemente sugerido usar adecuadamente la variable z para realizar la primera ramificación del problema.

P2. (a) La función de entropía $H(x) = -\sum_{i=1}^n x_i \ln x_i$ (definida para $x_i > 0$) tiene diversas aplicaciones en ingeniería, que van desde la codificación de mensajes hasta la compresión de datos. Es por esto que en esta pregunta se estudiará el comportamiento de esta función, desde el punto de vista de la optimización.

(i) Estudie el máximo de la función de entropía H en su dominio. Justifique que efectivamente se obtiene un máximo. Si ahora se redefine H para $x_i \geq 0$ considerando su comportamiento límite en 0 dado por $0 \ln 0 = 0$. Verifique que el máximo sigue siendo el mismo.

(ii) En la *teoría de la información* la función de entropía H representa una medida de incertidumbre asociada a una distribución de probabilidad $x = (x_i)_{i=1}^n$ de un mensaje emitido, es decir, la probabilidad de que un mensaje m sea igual a $i \in \{1, \dots, n\}$ es x_i (i.e. $\mathbb{P}(m = i) = x_i$). Así, entre mayor sea el valor de $H(x)$ mayor será la incertidumbre asociada a la distribución $x = (x_i)_{i=1}^n$. Muestre que la distribución de probabilidad que tiene máxima incertidumbre viene dada por $x_i = 1/n$ para todo $i = 1, \dots, n$. ¿Es lógica esta respuesta?

Indicación: Recuerde que toda distribución de probabilidad $x = (x_i)_{i=1}^n$ satisface que $\sum_{i=1}^n x_i = 1$ y $x_i > 0$ (pues gracias a la parte (i) esta última condición reemplaza a $x_i \geq 0$).

(iii) Para otras aplicaciones en la teoría de la información, la variable x debe satisfacer las condiciones:

$$0 < x_i \leq 1, \text{ para todo } i \in \{1, \dots, n\} \quad \text{y} \quad x_j = 1/n, \text{ para un y sólo un índice } j \in \{1, \dots, n\}.$$

Calcule el cono tangente asociado a estas restricciones en un $\bar{x}$ (arbitrario) que las satisfaga. Encuentre el punto crítico correspondiente a maximizar H sobre estas restricciones. ¿Podemos concluir que el punto obtenido es un máximo del problema? Comente.

Indicación: Dibuje el conjunto factible y resuelva la pregunta para el caso $n = 2$. El caso general donde n puede ser cualquier natural es análogo y sólo es necesario justificarlo correctamente.

Tiempo: 2 horas.
Sin calculadora. Sin consultas.