Probabilidades y Estadística (C)
Trabajo Práctico
1er. cuatrimestre de 2006

Algoritmos heurísticos para el problema de asignación

Un grupo de n personas debe ser asignado a un conjunto de n trabajos, de manera que cada persona sea asignada exactamente a un trabajo y cada trabajo a exactamente una persona. Se tiene un conjunto de n^2 valores $C_{i,j}$, $i, j = 1, \ldots, n$ que representan el costo $C_{i,j}$ de asignar a la persona i el trabajo j.

El problema clásico de asignación es el de determinar qué trabajo debe ser asignado a cada persona de manera de minimizar el costo total, es decir la suma de los n costos en los que se incurre.

1. ¿Cuántas formas hay de asignar los trabajos?

   Por lo tanto, calcular todos los posibles costos totales requiere un número muy grande de operaciones.

2. Supongamos que tenemos un software que computa todas las posibles asignaciones y para cada asignación calcula su costo y contamos con una computadora que tarda 10^{-20} seg. ($10^{-20} = 0{,}00000000000000000001$) por asignación. Calcular cuántos años tardaría en calcular todos los posibles costos para $n = 30$.

Los resultados del ítem 2 hacen que resulte imposible considerar todos los casos, por lo tanto introduciremos dos algoritmos heurísticos para tratar este problema.

Algoritmo Heurístico A

El primer algoritmo comienza asignando a la persona 1 el trabajo que resulta en el menor costo (para esta persona), es decir, la persona 1 es asignada al trabajo j_1 que satisface

$$C_{1,j_1} = \min_{1 \leq j \leq n} C_{1,j}.$$

Este trabajo es eliminado de la lista y a la persona 2 se le asigna el trabajo j_2 que resulta en el menor costo para esta persona (sin tener en cuenta el trabajo que ya fue asignado a la persona 1). Es decir

$$C_{2,j_2} = \min_{\substack{1 \leq j \leq n \\ j \neq j_1}} C_{2,j}.$$

De esta forma se continúa hasta que las n personas son asignadas.

Algoritmo Heurístico B

El segundo algoritmo es una versión más "global" del descripto anteriormente. Considera todos los n^2 costos y elige el par (i_1, j_1) para el cual $C_{i,j}$ es mínimo, es decir

$$C_{i_1,j_1} = \min_{\substack{1 \leq i \leq n \\ 1 \leq j \leq n}} C_{i,j}.$$

Se le asigna entonces el trabajo j_1 a la persona i_1 y se eliminan todos los costos $C_{i,j}$ que involucran al trabajo j_1 o a la persona i_1 (quedan $(n-1)^2$ costos) y se continua de la misma forma: se elige el par (i_2, j_2) que satisface

$$C_{i_2,j_2} = \min_{\substack{1 \leq i,\ j \leq n \\ i \neq i_1 \\ j \neq j_1}} C_{i,j},$$

etc. Es decir, que en cada etapa se elige la persona y el trabajo que minimizan el costo entre todos los no-asignados.

3. Exhibir para cada heurística un ejemplo en el que la solución obtenida no sea óptima.

4. Si suponemos que los costos $C_{i,j}$ son n^2 variables aleatorias independientes con distribución exponencial de parámetro 1. ¿Cuál de los dos algoritmos tiene menor valor esperado para el costo total?

   Si bien esta pregunta no es de fácil respuesta, podría ser respondida a través de un estudio de simulación. Una simulación es una recreación matemática de un proceso aleatorio. Mediante el uso de simulaciones se estudian procesos físicos, de ingeniería, económicos o biológicos. Algunos problemas complejos son analizados en una primera instancia mediante una simulación de manera de tener una aproximación al fenómeno en estudio.

   Esta recreación se realiza haciendo experimentos con muestreos pseudo–aleatorios en una computadora. El proceso bajo estudio se repite k veces y luego se analiza el comportamiento del proceso a lo largo de las k repeticiones (tamaño de la simulación).

   Responderemos la pregunta planteada mediante simulaciones del experimento.

   a) Hacer estas simulaciones para $n = 10$ y $n = 30$, pero escribiendo los programas de manera general (es decir, para n arbitrario). Aproximar el valor esperado para el costo total de cada algoritmo mediante una simulación de tamaño $k = 1000$.

   En función del resultado obtenido, ¿alguno de los dos métodos es preferible si tomamos como criterio de comparación que tenga el menor valor esperado? ¿Qué teorema usamos para asegurar que el valor calculado mediante simulaciones se asemeja al verdadero valor esperado?

   b) Escribiendo el costo óptimo como suma de variables aleatorias, calcular para la heurística A de manera teórica el costo esperado. Verificar si la aproximación hallada en a) es buena.

   c) Si se quisiera ver cómo varía el costo óptimo aproximado basado en k repeticiones, se debería repetir la simulación hecha en a) un número L de veces y hacer, por ejemplo, gráficos de lo obtenido. Para

   $$n = 30, \quad k = 100, \quad L = 50,$$

   graficar boxplots de los resultados obtenidos mediante ambas heurísticas. ¿Se parece el gráfico al comportamiento teórico que se esperaría? ¿Qué resultado teórico está usando?

   d) Variando k, graficar en un mismo gráfico los boxplots de los costos óptimos para ambos algoritmos para $n = 10$, $L = 50$ y valores de $k = 30, 100, 500$.