

EL PROBLEMA DEL MULTICUBRIMIENTO: UNA APLICACIÓN PARA LA SELECCIÓN DE PARADAS EN LA RED DE TRANSPORTE DE LA CIUDAD DE MÉXICO

MIGUEL ANGEL GUTIÉRREZ ANDRADE*

SERGIO GERARDO DE LOS COBOS SILVA**

BLANCA ROSA PÉREZ SALVADOR*** JOHN GODDARD CLOSE****

Recibido (versión revisada): 13 Jun 2003

Resumen

En este artículo se desarrolla un algoritmo heurístico y su correspondiente implementación para resolver un problema de muestreo en la red de rutas de transporte urbano de la Ciudad de México. El problema consiste en la selección de al menos 2 puntos (paradas de la ruta) en cada una de las 236 rutas en el estudio con un total de 8390 paradas. El problema anterior, se plantea como un problema de multicubrimiento (multicover problem) con 236 restricciones y 8390 variables binarias. Este problema es un problema NP-duro, por lo que se implementó un algoritmo heurístico para obtener los puntos de muestreo.

Palabras clave: Problema de multicubrimiento, métodos heurísticos, algoritmos glotones, optimización combinatoria.

Abstract

In this paper an heuristic algorithm is developed. Its implementation is developed as well, in order to solve a sampling problem in the urban transportation network in

*Departamento de Sistemas, Universidad Autónoma Metropolitana – Azcapotzalco, Av. San Pablo No. 180, Col. Reynosa Tamaulipas, Del. Azcapotzalco, México, D.F., C.P. 02200, México; Fax: +(52)53.94.45.34. E-Mail: gama@correo.azc.uam.mx.

**Departamento de Ingeniería Eléctrica, Universidad Autónoma Metropolitana – Iztapalapa, Av. Michoacán y La Purísima s/n, Col. Vicentina, Del. Iztapalapa, México, D.F., C.P. 09340, México; Fax: +(52)58.04.46.40. E-mail: cobos@xanum.uam.mx.

***Departamento de Matemáticas, UAM-I, misma dirección que de los Cobos. E-Mail: psbr@xanum.uam.mx

****Misma dirección que de los Cobos. E-Mail: jgc@xanum.uam.mx

Mexico City. The problem consist of the selection of at least 2 points (stops) on each of the 236 routes in the study comprising a total of 8390 stops. This problem is presented as a multicover problem subject to 236 restrictions and 8390 binary variables. Given that this an NP-hard problem, it was implemented an heuristic algorithm to determine the sampling points.

Keywords: Multicover problem, heuristic methods, greedy algorithms, combinatorial optimization.

Mathematics Subject Classification: 90C27

1. Introducción

El Gobierno de la Ciudad de México llevó a cabo un proyecto denominado “Estructuración Física de la Red de Transporte de Superficie Autobuses y Trolebuses”, cuyo objetivo principal es reestructurar el servicio de transporte público que se presta en la Ciudad de México a través de autobuses y trolebuses.

Los objetivos específicos del proyecto son:

- Reasignación y diseño de nuevas rutas.
- Coordinación de modos de transporte.
- Definición de flota vehicular.
- Programación y control del servicio.

Para alcanzar los objetivos anteriormente mencionados, se propusieron 3 etapas y en cada una de estas etapas se definieron actividades específicas.

La Etapa 1 se denominó: Planeación de la red de transporte de la Ciudad de México y se definieron las siguientes actividades:

- Recopilación y validación de la información.
- Investigación de la oferta, demanda y recaudación.
- Investigación de ingresos y costos.
- Diagnóstico de rutas.
- Estructura organizacional del transporte.

La Etapa 2 se denominó: Modernización de la red y programación de la operación y se definieron las siguientes actividades:

- Ajuste del trazo y diseño de nuevas rutas.
- Programación de la operación.
- Sistema de mantenimiento de autobuses e infraestructura requerida.

- Programa de implementación.

La Etapa 3 se denominó: Monitoreo y se definieron las siguientes actividades:

- Armonizar resultados del modelo con la realidad.
- Actualización de la información de campo.
- Utilización de modelos para retroalimentación.
- Revisión periódica y ajuste de rutas.
- Equipamiento que permita monitorear el parque vehicular.

Para obtener la información de campo que se requiere en este tipo de proyectos se realizaron las siguientes actividades:

- Realizar aforos en algunos puntos de la red, para obtener: periodos de máxima demanda, variación semanal y composición del parque vehicular.
- Obtener Frecuencias de paso y cargas, es decir, frecuencia de paso en cada una de las rutas, intervalos de paso, horas de máxima demanda, volumen de pasajeros (tendencias)
- Obtener información de ascensos y descensos, polígonos de carga, ocupación promedio de la unidad, ascensos promedio por viaje, sección de carga máxima, paradas importantes, volumen de pasajeros atendidos, tiempos de parada.
- Hacer una encuesta de sondeo de opinión en donde se obtenga información tal como la matriz de origen-destino, cadena modal, base de información para alimentar modelos, opinión de los usuarios por ruta, características socioeconómicas del usuario por zona.

La red de transporte de la Ciudad de México está cubierta por 163 rutas de las cuales 138 rutas tienen recorridos de ida y vuelta y 25 rutas son circuitos. En la tabla de abajo se da un resumen de las rutas. En la primera columna aparece el nombre de la empresa que opera la ruta, en la segunda columna se da el número de rutas que opera la empresa, en la tercera columna aparece el número promedio de corridas diarias que se proporciona la empresa y en la cuarta columna se da el número promedio por día de pasajeros que mueve cada empresa.

EMPRESA	No. de RUTAS	No. de CORRIDAS	No. de PASAJEROS
Consejo de Incautación	83	6,127	358,593
STE Trolebuses	17	5,736	207,185
STE Articulados	8	605	95,415
17 de marzo	12	1,689	177,042
Siglo Nuevo (Bombas)	7	748	77,952
Siglo Nuevo (Santa Marta)	11	2,395	94,530
Transportes G	25	4,764	358,545
TOTAL	163	22,065	1,367,262

Para obtener parte de la información mencionada anteriormente como es: la encuesta origen-destino, encuesta de sondeo de opinión y el monitoreo de las rutas en diferentes puntos de la red, un problema importante en esta red de transporte urbano, es la localización de los puntos en donde se van a colocar dispositivos para su monitoreo, o para cuantificar ciertos parámetros como la frecuencia de paso, carga, y otras características de las rutas de transporte. En este estudio se decidió, por cuestiones técnicas, seleccionar al menos 2 paradas de cada una de las rutas con el objeto de monitorear la ruta y cuantificar las frecuencias de paso, corridas, tiempos de recorrido, sondeos de opinión, y evasión de pago, de la red de transporte urbano, para posteriormente hacer inferencia, de los parámetros antes mencionados, en la población.

2. Modelo

El problema de seleccionar dos o más puntos de la red para cuantificar frecuencia de paso y carga se puede plantear de la siguiente manera: suponga que se tienen m rutas de transporte sobre la red que se desean monitorear y existen n puntos (paradas de las rutas) en donde se pueden monitorear las rutas. Se desea seleccionar un subconjunto de estos puntos con la característica que haya al menos dos puntos de este subconjunto que pertenezcan a cada una de las m rutas. También se desea que este subconjunto de puntos sea, en cuanto al número total de puntos, lo más chico que se pueda para que el monitoreo de la red sea lo más barato posible. Este problema se puede plantear en términos de programación lineal entera de la siguiente manera. Supongamos que se denota al conjunto de puntos por $\{1, 2, \dots, n\}$, se definen las siguientes variables binarias:

$$x_j = \begin{cases} 1 & \text{si el punto } j \text{ se selecciona para efectuar los aforos, } j = 1, \dots, n \\ 0 & \text{en otro caso} \end{cases}$$

Así, el problema se puede plantear como:

$$\begin{aligned} \text{mín } z &= x_1 + x_2 + \dots + x_n \\ \text{Sujeto a:} & \sum_{O_i} x_j \geq 2 \quad \text{para toda } i = 1, 2, \dots, m \\ & x_j = 0 \text{ o } 1 \quad \text{para toda } j = 1, 2, \dots, n \end{aligned} \quad (1)$$

donde

O_i es el conjunto de puntos que están sobre la ruta i para $i = 1, \dots, m$

A este tipo de modelo se le conoce en la literatura como el Problema de Multicubrimiento (*multicover problem*), vea Hochbaum (1997). La función objetivo minimiza el número de puntos a ser seleccionados para efectuar los aforos y esta función objetivo esta sujeta a que se seleccionen por lo menos dos puntos de cada ruta, esto es, hay una restricción para cada una de las m rutas del tipo mayor o igual a 2, finalmente se restringe a las variables a ser binarias (0 ó 1).

El problema dual asociado con el sistema (1) es el siguiente

$$\begin{aligned} \text{máx } w &= 2y_1 + 2y_2 + \cdots + 2y_m \\ \text{Sujeto a:} & \sum_{T_j} y_i \leq 1 \quad \text{para toda } j = 1, 2, \dots, n \\ & y_i \geq 0 \quad \text{para toda } i = 1, 2, \dots, m \end{aligned}$$

donde

T_j es el conjunto de rutas que pasan por el punto j para $j = 1, \dots, n$

si las variables y_i se restringen a ser binarias, entonces este problema se conoce en la literatura como el problema de empaquetamiento de un conjunto (*set packing problem*).

Para una red con pocas rutas el problema resulta fácil de resolver, pero cuando el número de rutas y el número de puntos se incrementa, la solución del problema ya no resulta tan trivial, es más, resulta ser un problema difícil de resolver. Para resolver el problema del sistema (1), se hicieron pruebas con ejemplos de redes de transporte de alrededor de 100 rutas y 500 paradas. En la solución de las instancias propuestas, se usaron los paquetes LINDO (*Linear, INteractive, and Discrete Optimizer*), GAMS (*General Algebraic Modeling System*) y CPLEX 6.5 y no convergieron en un tiempo razonable.

3. Heurística para el problema de multicubrimiento

El problema de multicubrimiento se puede formular como un problema entero 0-1 puro como sigue:

$$\begin{aligned} \text{mín } z &= cx \\ \text{Sujeto a: } Ax &\geq b \\ x_j &= 0 \text{ o } 1 \text{ para toda } j \end{aligned} \tag{2}$$

donde c es un vector fila de n componentes, $A = (a_{ij})$ es una matriz de orden $m \times n$ con todas sus entradas iguales a 0 o 1 y b es un vector columna en R^m de entradas enteras mayores a cero.

En el sistema (2) una variable x_j se dice que *cubre* la i -ésima restricción si x_j aparece con un coeficiente +1 en esta restricción. Si x_j cubre la i -ésima restricción, cualquier vector x con entradas 0-1 en donde la variable $x_j = 1$, contribuye en una unidad para satisfacer la restricción y cada restricción requiere tener al menos b_i variables iguales a 1 para que la restricción se satisfaga, donde b_i es la i -ésima entrada del vector b . Algunos resultados que ayudan a fijar los valores de las variables en 0 o en 1, y eliminar algunas restricciones, y por lo tanto reducir el problema a un problema equivalente pero con menos restricciones, menos variables o valores menores en el lado derecho son:

RESULTADO 1 Si $c \leq 0$, una solución óptima para el sistema (2) es tomar $x = e_n$, el vector en R^n con todas las entradas iguales a 1. Terminar.

RESULTADO 2 En el sistema (2) suponga que $c \not\leq 0$. Si j es tal que $c_j \leq 0$ fijamos la correspondiente variable x_j en 1 y restamos 1 a los valores b_i del lado derecho en cada

restricción i donde aparezca x_j , si después de restar uno, algún lado derecho se hace cero, se elimina la restricción. Si no quedan restricciones, fijamos el resto de las variables en 0, esto da una solución óptima al problema.

RESULTADO 3 Si j es tal que $c_j > 0$ y la variable x_j no aparece en las restricciones restantes con un coeficiente $+1$, fije la variable x_j en 0.

Aplique los resultados 2 y 3 tantas veces como sea posible para reducir el problema. Después de hacer este proceso, el problema se reducirá y tendrá la misma forma de (2), de tal manera que toda variable que aparezca tendrá un coeficiente positivo en la función objetivo. Un método glotón (greedy) se le aplicará a este problema reducido, y consiste en ejecutar el siguiente paso general, de manera iterativa. Aquí una variable libre se refiere a una variable cuyo valor no ha sido fijado en 1 o en 0.

PASO GENERAL En el problema restante, para cada variable libre x_j , sea d_j el número de restricciones cubiertas por x_j . c_j/d_j se puede interpretar como el costo por restricción cubierta, asociada con la variable libre x_j en este paso. Encuentre una variable libre x_r con el menor costo asociado por restricción cubierta en el problema reducido. Esto es, tome r tal que, $c_r/d_r = \min\{c_j/d_j : j \text{ tal que } x_j \text{ es una variable libre}\}$. Fije x_r en 1, reste 1 a todos los lados derechos de las restricciones en donde la variable x_r aparezca y elimine todas las restricciones cuyo lado derecho se haya hecho igual a 0. Si ya no quedan restricciones fije el resto de las variables libres en cero y termine. De otra manera aplique el resultado 3 al problema restante tantas veces como sea necesario y repita el paso general.

4. Ejemplo

Suponga que se tiene una red de transporte con 5 rutas. La primera ruta pasa por las paradas etiquetadas por $\{1,2,6,7,11,12\}$, la segunda ruta pasa por las paradas etiquetadas por $\{3,4,6,7,9,10\}$, la tercera ruta pasa por las paradas etiquetadas por $\{5,6,7,8,10,11\}$, la cuarta ruta pasa por las paradas etiquetadas por $\{1,2,3,7,8\}$ y la quinta ruta pasa por las paradas etiquetadas por $\{3,4,7,9,10,11\}$. De acuerdo a lo dicho anteriormente, el problema lineal entero 0-1 queda como:

$$\begin{aligned}
 \text{mín } z = & \quad x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + \\
 & \quad x_7 + x_8 + x_9 + x_{10} + x_{11} + x_{12} \\
 \text{Sujeto a:} & \\
 \text{Ruta 1} & \quad x_1 + x_2 + x_6 + x_7 + x_{11} + x_{12} \geq 2 \\
 \text{Ruta 2} & \quad x_3 + x_4 + x_6 + x_7 + x_9 + x_{10} \geq 2 \\
 \text{Ruta 3} & \quad x_5 + x_6 + x_7 + x_8 + x_{10} + x_{11} \geq 2 \\
 \text{Ruta 4} & \quad x_1 + x_2 + x_3 + x_7 + x_8 \geq 2 \\
 \text{Ruta 5} & \quad x_3 + x_4 + x_7 + x_9 + x_{10} + x_{11} \geq 2 \\
 & \quad x_j = 0 \text{ o } 1 \text{ para toda } j = 1, 2, \dots, 12
 \end{aligned} \tag{3}$$

Ahora aplicaremos el algoritmo glotón al problema dado en (3). Los resultados 1 y 2 no

se aplican para este problema ya que no existe ningún coeficiente menor o igual a cero en la función objetivo. Al inicio tampoco el resultado 3 se aplica, ya que todas las variables aparecen con +1 en al menos una de las restricciones. Ahora apliquemos el paso general al problema dado en (3). En la tabla de abajo aparece en la primera fila los nombres de las variables de decisión, en las siguientes cinco filas aparece un 1 en la entrada (i, j) si la variable de decisión x_j aparece con un +1 en la i -ésima restricción y en blanco en caso contrario. En la última fila aparecen los valores de las d_j .

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	
1	1				1	1				1	1	≥ 2
		1	1		1	1		1	1			≥ 2
				1	1	1	1		1	1		≥ 2
1	1	1				1	1					≥ 2
		1	1			1		1	1	1		≥ 2
2	2	3	2	1	2	5	2	2	3	3	1	d_j

Para este ejemplo los valores c_j todos son iguales a 1 y por lo tanto no los reportaremos en la tabla. El criterio para fijar en 1 las variables libres, es equivalente a seleccionar la variable libre con el mayor d_j que en este caso es x_7 . De modo que fijamos $x_7 = 1$, y a los lados derechos de todas las restricciones les restamos 1 (ya que x_7 aparece en todas las restricciones). El problema resultante queda como:

x_1	x_2	x_3	x_4	x_5	x_6	x_8	x_9	x_{10}	x_{11}	x_{12}	
1	1				1				1	1	≥ 1
		1	1		1		1	1			≥ 1
				1	1	1		1	1		≥ 1
1	1	1				1					≥ 1
		1	1				1	1	1		≥ 1
2	2	3	2	1	2	2	2	3	3	1	d_j

Este problema tiene las mismas restricciones que el original con la diferencia que ahora el lado derecho aparece reducido en una unidad y la variable x_7 que fijamos en 1 ya no está considerada. En esta tabla las variables libres con mayor valor d_j son: x_3, x_6, x_{10}, x_{11} . De modo que podemos seleccionar cualquiera de ellas para fijarla en 1. Tomemos en forma arbitraria $x_3 = 1$, y a los lados derechos de todas las restricciones en donde aparece x_3 se les resta 1. Los lados derechos de las restricciones 2,4 y 5 se anulan y de acuerdo al paso general, se eliminan. El problema resultante queda como:

x_1	x_2	x_4	x_5	x_6	x_8	x_9	x_{10}	x_{11}	x_{12}	
1	1			1				1	1	≥ 1
			1	1	1		1	1		≥ 1
1	1	0	1	2	1	0	1	2	1	d_j

Aquí observamos que las variables x_4 y x_9 no aparecen en las restricciones restantes, si aplicamos el resultado 3 entonces fijamos $x_4 = x_9 = 0$, y ahora nos queda el problema:

$$\begin{array}{cccccccc|c}
 x_1 & x_2 & x_5 & x_6 & x_8 & x_{10} & x_{11} & x_{12} & \\
 \hline
 1 & 1 & & 1 & & & 1 & 1 & \geq 1 \\
 & & 1 & 1 & 1 & 1 & 1 & & \geq 1 \\
 \hline
 1 & 1 & 1 & 2 & 1 & 1 & 2 & 1 & d_j
 \end{array}$$

Las variables libres con mayor valor d_j ahora son x_6 y x_{11} . De modo que podemos seleccionar cualquiera de ellas para fijarla en 1. Tomemos $x_6 = 1$, y a los lados derechos de todas las restricciones en donde aparece x_6 se les resta 1. Los lados derechos de las restricciones 1 y 3 se anulan y por lo tanto se eliminan. Como ya no queda ninguna restricción que satisfacer, fijamos los valores de las variables restantes en cero, por lo tanto, $x_1 = x_2 = x_5 = x_8 = x_{10} = x_{11} = x_{12} = 0$. Finalmente, obtenemos una solución al problema original igual a $x_3 = x_6 = x_7 = 1$ y el resto de las variables iguales a 0. Para este caso, la solución encontrada es una solución óptima pero, en general, el algoritmo no necesariamente encuentra la solución óptima al problema.

5. Conclusiones

- El problema de multicubrimiento planteado para la selección de paradas en la red de transporte de la Ciudad de México es un problema clasificado como NP-duro y por lo tanto no se conocen algoritmos polinomiales que lo resuelvan.
- Para instancias de más de 100 variables binarias hay problemas de convergencia con paquetes comerciales tales como LINDO (*Linear, INteractive, and Discrete Optimizer*), GAMS (*General Algebraic Modeling System*) y CPLEX 6.5.
- Se desarrolló e implementó un método glotón para obtener una solución “buena” al problema planteado.
- En las pruebas que se hicieron, en donde se conocía la solución óptima, el algoritmo glotón proporcionó soluciones óptimas.
- El algoritmo desarrollado tiene la característica de tomar puntos por donde pasan muchas rutas, esto permite que rutas importantes sean monitoreadas en más de dos puntos.
- Para fines prácticos, en general, las paradas seleccionadas por el algoritmo son muy transitadas y de fácil acceso.
- La técnica permite con facilidad modificar el criterio de seleccionar dos o más puntos de cada ruta por cualquier otro número mayor.

- En caso de empate, al seleccionar la variable que se va a fijar en uno, se puede proporcionar un criterio adicional para romper empates asignándole a cada punto algunas características adicionales.
- Para el problema de transporte se seleccionaron 348 paradas de las 8390 posibles para monitorear la red de transporte, es decir, se está seleccionando un promedio de 1.47 puntos por ruta.

Referencias

- [1] Chvátal, V. (1979) “A greedy heuristic for the set-covering problem”, *Mathematical of Operations Research* **4**(3): 233–235.
- [2] Hochbaum, D.S. (1997) *Approximation Algorithms for NP-Hard Problems*, PWS Publishing Company, Boston, USA.
- [3] Murty, K.G. (1995) *Operations Research Deterministic Optimization Models*. Prentice Hall, New Jersey, USA.
- [4] Pisinger, D. (2002) “Heuristics for the container loading problem”, *European Journal of Operational Research* **141**: 382–392.
- [5] Wenxun, X. (2002) “A bin packing problem with over-sized items”, *Operations Research Letters*, **30**(3): 832–888.
- [6] Willoughby, K.A. (2002) “A mathematical programming analysis of public transit systems”, *Omega* **30**: 137–142.