

Modelos y algoritmos en la solución de un problema real de agrupamiento y planificación de visitas a clientes.

María Gulnara Baldoquin de la Peña

Facultad de Ingeniería, Pontificia Universidad Javeriana Cali
Calle 18 N° 118-250 Vía a Pance Cali. Colombia
mgulnara@javerianacali.edu.co

Rodrigo Linfati

Department of Electronics, Computer Sciences and Systems, University of Bologna
Viale Risorgimento 2, 40136 Bologna, Italy
rodrigo.linfati@unibo.it

Annalay Escalera Fariñas

Centro de Estudios de Matematica, ISPJAE
Calle 127, Marianao, La Habana, Cuba
annalay@cemat.cujae.edu.cu

Resumen.

El trabajo es parte de un problema logístico real y complejo que surge en una empresa mixta cuyo objetivo fundamental es la comercialización de su producto en el mercado nacional. El problema combina decisiones estratégicas y tácticas, de agrupamiento, planificación mensual de visitas a clientes y ruteo. Se muestra la metodología propuesta en su solución, subdividiendo el problema en fases, pero teniendo en cuenta en cada fase elementos de otras fases. Se hace un resumen de cómo se resuelve el agrupamiento de clientes, centrando el trabajo en el modelo y método de solución propuesto para la fase de planificación de visitas mensual a clientes. El modelo es no lineal involucrando variables binarias y continuas, y se resuelve por un método heurístico. Los experimentos computacionales muestran que el método de solución propuesto obtiene buenos resultados para problemas de la vida real e instancias de la literatura adaptadas al problema abordado.

Palabras claves: logística, planificación, heurística.

Área principal: Logística y Transporte, OR en la industria, Optimización Combinatoria.

Abstract.

The paper is part of a real and complex logistical problem that arises in a mixed company whose fundamental objective is the commercialization of its product in the national market. The problem combines strategic and tactical decisions, of *cluster*, monthly planning of visits to clients and routing. The methodology proposed for its solution is shown, subdividing the problem in phases, but considering some relations between phases. A summary is made of how the clustering phase was solved, centering the work in the model and solution method proposed for the phase of monthly scheduling of visits to clients. The model is nonlinear, involving binary and continuous variables, and it is solved with a heuristic method. The computational experiments show that the proposed solution method performed very well over real life and literature instances.

Keywords: logistic, scheduling, heuristic.

Main Area: L & T - Logistics and Transport, IND – OR in Industry, OC - Combinatorial Optimization.

Introducción.

El trabajo presentado es parte de un problema logístico real y complejo que surge en una empresa con capitales públicos y privados cuyo objetivo fundamental es la comercialización de su producto en el mercado nacional.

El aporte fundamental, desde el punto de vista práctico en la solución del problema real presentado, está dirigido a la mejora en la cadena de suministro a través de una de sus dos principales actividades, la logística, donde se combinan decisiones estratégicas y tácticas. El problema general combina la partición de clientes en determinados grupos con la planificación de las visitas a dichos clientes sobre un horizonte de tiempo dado (decisiones estratégicas) y la búsqueda de rutas eficientes para la realización de las visitas a los clientes (decisiones tácticas). Desde el punto de vista teórico, este problema involucra subproblemas de agrupamiento, planificación y ruteo de vehículos, interdependientes, y con determinadas complejidades impuestas por restricciones, que no son considerados por modelos clásicos encontrados en la literatura.

Resolver el problema como un todo, involucrando agrupamiento, localización y ruteo, es muy complejo. La literatura es amplia en la solución de problemas integrados entre los que pueden mencionarse localización-ruteo (Zhao et al. (2001), Nagy and Salhi (2007)), asignación-ruteo (Cortés and García, (2007)), agrupamiento-ruteo (Deng and Bard (2010), González and González (2007)), planificación de tripulaciones (Freling et al. (2003)), problemas de planificación y ruteo (Korsvik and Fagerholt (2010), Averbakh and Berman (1999)), inventario-localización de facilidades–modos de transporte (Jayaraman (1998)). La literatura también registra ejemplos de problemas logísticos complejos que son divididos en fases por su nivel de complejidad. Pueden mencionarse el problema de diseño de redes de tránsito urbana (*UTNDP*, por sus siglas en inglés) divididos en dos componentes principales: ruteo y planificación (Chakroborty (2003)); el problema de planificación de tripulaciones, donde usualmente se reduce a 3 fases (Caprara et al. (2001)). En (Bar-Ilan et al. (1993)) se presenta diferentes métodos para la selección de centros de distribución y la asignación de clientes a estos. Un estudio de las diferentes variantes del problema de asignación y sus métodos de solución se puede encontrar en (Pentico (2007)). El modelo propuesto es Entero Mixto No Lineal No Convexo, un estudio sobre los métodos de solución para este tipo de modelos se encuentra en (Letchford and Burer (2012)).

En este trabajo se hace un breve resumen de la metodología de solución del problema completo, dividiéndolo en fases, así como del método de solución de la Fase 1. El aporte fundamental está en el modelo y método de solución del subproblema planteado en la Fase 2. La solución de la Fase 3 es objeto de un trabajo posterior. Se plantea un diseño de experimentos para validar el método de solución planteado, utilizando información del problema real e instancias (ejemplos) de la literatura válidas para este tipo de problemas.

La estructura del problema de la vida real es la siguiente: Existe un conjunto de puntos de venta del producto que deben ser visitados todos en el mes con una frecuencia definida por su volumen de ventas, por uno de los 9 camiones con los que cuenta la empresa, desde un mismo depósito central al cual deben regresar al finalizar su jornada. Cada punto debe ser siempre visitado por el mismo camión. Existen puntos que deben ser visitados el mismo día por el mismo camión, esto ocurre cuando se visita al menos una tienda que está asociada a una llamada “casa matriz”. Existen cuatro tipos de frecuencia de visita, dependiendo del volumen de venta a cada punto: Semanal (manteniendo el mismo día en las 4 semanas en el mes), Bisemanal (siempre Lu-Ju ó Ma-Vi), Quincenal (el mismo día de la semana en las Semanas 1 y 3 ó Semanas 2 y 4), Mensual (solo una vez al mes). Solo se distribuye de lunes a viernes. El mes se considera de 4 semanas con 20 días de entrega.

Existen en los puntos de venta diferentes tiempo de estadía, dependiendo del tipo de visita en dicho punto, pero no del volumen de ventas. El tiempo de permanencia de un camión en los puntos de venta no es constante, ni depende de la carga entregada, sino del tipo de visita que se realiza en el punto. Aunque existe una capacidad limitada de los camiones, la restricción actual fundamental es de limitación de tiempo total disponible de recorrido del camión, no de su capacidad, por tanto la capacidad de los camiones es en general superior a la carga que llevan para la cantidad máxima de puntos de venta que pueden visitar en un día, teniendo en cuenta todos los aspectos del ruteo que no se detallarán en este trabajo.

Metodología propuesta.

El problema se plantea resolverlo de forma heurística, involucrando 3 fases fundamentales, en el orden que se mencionan:

1. Problema de sectorización de clientes, basado en el uso de técnicas de *clustering*.
2. Problema de planificación de visitas a clientes que se encuentran en un mismo *cluster* en un periodo determinado, actualmente de un mes.
3. Ruteo de vehículos con ventanas de tiempo y otras restricciones adicionales, para cada conjunto de clientes a visitar cada día por un mismo camión (vendedor).

Estas 3 fases no se resuelven de forma totalmente independientes. En la Fase 1 se tienen en cuenta restricciones que faciliten en la fase posterior de planificación de visitas a los clientes una solución adecuada, tratando de balancear el volumen de visitas mensuales a realizar en cada grupo formado y que contenga en el mismo *cluster* aquellos clientes con igual frecuencia de visitas que deben ser visitados juntos. Ello conlleva la construcción de *clusters* con restricciones, fundamentalmente de capacidad. De igual forma en la Fase 2 se tienen en cuenta aspectos para lograr rutas eficientes en la Fase 3, buscando minimizar la distancia total de los puntos visitados cada día de su centroide.

Fase 1: El problema de agrupamiento de clientes.

Para dar solución al problema de la conformación de las zonas de distribución con restricciones de capacidad se decidió seguir una filosofía similar a la propuesta por Pacheco y Beltrán en (Pacheco and Beltrán (2001)) y por Éric Taillard en (Taillard (2003)). Los mismos son adaptaciones del conocido método *k-means* con el uso de centroides, lo que le proporciona una clara interpretación física al problema abordado (agrupamiento en zonas geográficas), aportando el centro geométrico de cada *cluster*.

Los algoritmos diseñados constan de dos fases que consisten básicamente en: 1) seleccionar una cantidad k de puntos iniciales (semillas), de manera que se encuentren suficientemente alejados entre sí y que serán considerados los centroides iniciales de los k *clusters* a construir y 2) asignar uno a la vez, el resto de los puntos (clientes) a algún *cluster* o zona, siguiendo la filosofía de los algoritmos *k-medias* y adaptándolos de manera que las restricciones de capacidad sean tenidas en cuenta. La capacidad de un *cluster* es la cantidad máxima de visitas mensual a realizar en el mismo. Los clientes que deben ser visitados juntos se sustituyen en esta fase por uno solo, con una “capacidad” igual a la suma de todos ellos y una posición que es el centroide de todos esos puntos.

Con el objetivo de obtener diferentes métodos de solución, las dos fases descritas anteriormente fueron diseñadas de manera independiente, dando la posibilidad de combinarlas posteriormente. Por cada fase fueron implementados dos métodos distintos, de esta manera se obtienen cuatro métodos para la creación de los *clusters* o zonas de distribución.

Se realizó un diseño de experimentos para validar el desempeño de los mismos, en cuanto a la calidad de solución encontrada y tiempo de ejecución. Se tuvieron en cuenta para ello como parámetros que se variaron: número de agrupamientos a realizar, cantidad de clientes y estructura de distribución de los puntos ó clientes totales, en 3 clases:

- La clase de las instancias donde sus puntos están distribuidos uniformemente en la región en la cual se encuentran. Esta clase se denota con la letra *r*.
- La clase de las instancias donde sus puntos están concentrados en determinadas áreas de la región en la cual se encuentran. Esta clase se denota con la letra *c*.
- La clase de las instancias que es una combinación de las características de las dos clases anteriores. Esta clase se denota con la letra *rc*.

En la Figura 1 se muestra un ejemplo de instancias para diferentes clases

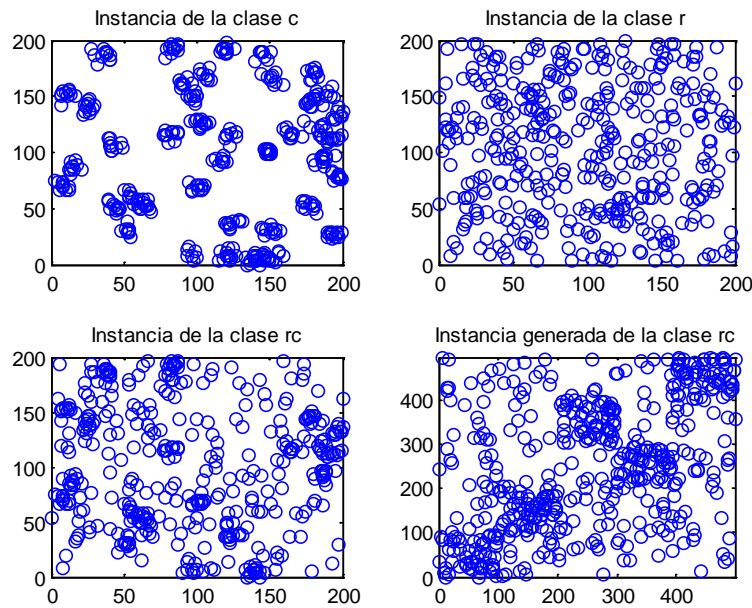


Figura 1: Representación gráfica de clientes que corresponden a instancias de las clases *c*, *r* y *rc*.

De acuerdo a los resultados obtenidos, una de las variantes de algoritmos no compite con las demás en cuanto a calidad de soluciones. Del resto se concluyó que ningún algoritmo es dominante sobre los demás teniendo en cuenta todos los elementos que pueden incidir simultáneamente en el desempeño de los algoritmos. Sin embargo, de acuerdo a las características particulares que presenten las instancias de un usuario y los clusters a conformar, los resultados del análisis realizado permiten sugerir el uso de uno de dos de los algoritmos propuestos (Escalera, (2011)).

Fase 2: Planificación de las visitas mensuales a clientes en un *cluster* dado.

- **Modelo matemático propuesto para la asignación de visitas mensuales a puntos de venta.**

Índices.

- j*: día
- f*: frecuencia
- l*: cliente
- k*: casa matriz

Conjuntos.

F : conjunto de frecuencias f , $F = \{(BS) \text{ Bisemanal}, (S) \text{ Semanal}, (Q) \text{ Quincenal}, (M) \text{ Mensual}\}$

L_f : Conjunto de clientes l con frecuencia $f \in F$

K : Conjunto de casas matrices k

A_k^f : Conjunto de clientes l con frecuencia f asociados a la casa matriz k , $A_k^f \subseteq L_f$

J_f : Conjunto de días del mes que corresponde a la frecuencia f , eliminando algunos que pueden deducirse de otros según restricciones impuestas.

$$J_f = \begin{cases} \{1,2\}, & f = \text{BiSemanal} \\ \{1,\dots,5\}, & f = \text{Semanal} \\ \{1,\dots,10\}, & f = \text{Quincenal} \\ \{1,\dots,20\}, & f = \text{Mensual} \end{cases}$$

Parámetros.

u_l = coordenada x para el cliente $l \in L_f$

v_l = coordenada y para el cliente $l \in L_f$

$$PROM = \text{ceil} \left(\frac{8 * |L_{BS}| + 4 * |L_S| + 2 * |L_Q| + 1 * |L_M|}{20} \right)$$

Funciones.

$$g_f(j) = \begin{cases} NULL, & f = \text{BiSemanal}, ((j-1) \bmod 5) = 2 \\ 1 + (((j-1) \bmod 5) \bmod 3), & f = \text{BiSemanal}, ((j-1) \bmod 5) \neq 2 \\ 1 + ((j-1) \bmod 5), & f = \text{Semanal} \\ 1 + ((j-1) \bmod 10), & f = \text{Quincenal} \\ j, & f = \text{Mensual} \end{cases}$$

Variables.

cx_j = coordenada x para el centroide correspondiente al día $j = 1, \dots, 20$

cy_j = coordenada y para el centroide correspondiente al día $j = 1, \dots, 20$

$$x_{lj}^f = \begin{cases} 1 & \text{si el cliente } l \text{ con frecuencia } f \text{ es visitado el día } j \\ 0 & \text{en otro caso} \end{cases}$$

Función objetivo.

$$\text{Min} \sum_{j=1}^{20} \sum_{f \in F} \sum_{l \in L_f} x_{lj}^f \sqrt{(u_l - cx_j)^2 + (v_l - cy_j)^2}$$

Restricciones.

$$\sum_{j \in J_f} x_{lj}^f = 1 \quad \forall l \in L_f \quad \forall f \in F \quad (1)$$

$$\sum_{l \in A_k^f} x_{lj}^f = |A_k^f| * x_{lj}^f \quad \forall l \in A_k^f \quad \forall j \in J_f \quad \forall f \in F \quad \forall k \in K \quad (2)$$

$$\sum_{f \in F} \sum_{l \in L_f} x_{lg_f(j)}^f \leq PROM \quad \forall f \in F \quad j = 1, \dots, 20 \quad (3)$$

$$\frac{\sum_{f \in F} \sum_{l \in L_f} u_l * x_{lg_f(j)}^f}{\sum_{f \in F} \sum_{l \in L_f} x_{lg_f(j)}^f} = cx_j \quad j = 1, \dots, 20 \quad (4)$$

$$\frac{\sum_{f \in F} \sum_{l \in L_f} v_l * x_{lg_f(j)}^f}{\sum_{f \in F} \sum_{l \in L_f} x_{lg_f(j)}^f} = cy_j \quad j = 1, \dots, 20 \quad (5)$$

$$x_{lj}^f \in \{0,1\} \quad \forall l \in L_f \quad \forall j \in J_f \quad \forall f \in F \quad (6)$$

$$cx_j, cy_j \geq 0 \quad \forall j \in J_f \quad \forall f \in F \quad (7)$$

El papel de la función $g_f(j)$ es poder relacionar las variables x_{lj}^f , para diferentes días j del mes y frecuencias diferentes de distribución, teniendo en cuenta que para reducir el número de las variables a utilizar no se definen 20 variables (días del mes a considerar) para cada tipo de frecuencias. Para la frecuencia BS solo se usan dos variables, para la frecuencia S son 5 variables, y para las frecuencias Q y M son 10 y 20 respectivamente.

La función objetivo busca minimizar la suma de las distancias desde los puntos a visitar cada día hasta su centroide (cx_j, cy_j) . Ello se tiene en cuenta para la solución del problema posterior de búsqueda de las mejores rutas de distribución diarias.

Las restricciones de tipo (1) tienen en cuenta que todo cliente debe tener una asignación en el mes de acuerdo a su frecuencia de visita. Ello no significa que se visita exactamente una sola vez. Así, por ejemplo, un cliente l con frecuencia $f = BS$ tiene dos posibilidades de asignación del conjunto $J_{BS} = \{1,2\}$. Si el valor es 1, significa que se visitará todas las semanas los lunes y miércoles, si es 2 se visitará todas las semanas los martes y viernes, por lo cual recibe 8 visitas en el mes. Un cliente l con frecuencia $f = S$ tiene cinco posibilidades de asignación del conjunto $J_S = \{1,2,3,4,5\}$. Si el valor es 1, significa que se visitará todas las semanas los lunes, si es 2 se visitará todas las semanas los martes, si es 3 se visitará todas las semanas los miércoles, etc.

Las restricciones de tipo (2) imponen que todos los clientes asociados a la misma casa matriz con igual frecuencia de visitas sean visitados los mismos días del mes.

Las restricciones de tipo (3) imponen balancear el número diario de visitas realizadas, según el número total de visitas programadas en el mes.

Las restricciones de tipo (4) y (5) van asignando los centroides para cada uno de los días del mes de acuerdo a las visitas programadas en esos días.

Las restricciones de tipo (6) y (7) definen el dominio de las variables.

- **Método de solución propuesto.**

El modelo es altamente no lineal, con variables de tipo binaria y continua. De acuerdo a los datos reales en la actualidad, para uno de sus *clusters*, el número de variables del problema sería de 40 variables continuas y aproximadamente 2700 variables binarias.

Se prevé que para el problema real, en los próximos años, esos valores pueden aumentar de forma sustancial, de acuerdo a un incremento previsto de puntos de ventas.

Por ello se propone un método heurístico de solución, que utilice como método subyacente un método exacto de solución de un Problema Lineal Entero (ILP, siglas en inglés).

Se comienza colocando en la coordenada (0,0) los centroides de todos los puntos a visitar para cada uno de los días del mes. Con los centroides fijados el modelo se convierte en un ILP que se resuelve usando algún *software* adecuado para ello como es ILOG CPLEX. Se sustituyen los valores x_{lj}^f de la solución obtenida en las ecuaciones (4) y (5) de los centroides. Se reactualiza el valor de la función objetivo $z(x)$ con los nuevos valores de los centroides. Si $z(x)$ es mejor que la mejor solución encontrada hasta el momento, se reactualiza la misma. Cuando se reconoce que se llega a un óptimo local entonces se trata de escapar del mismo y diversificar la búsqueda reactualizando los centroides con valores aleatorios y reiniciando otro proceso de intensificación de la búsqueda a partir de estos valores.

El mecanismo de escape debe ser a un punto "lejano" ya que si resulta similar se vuelve a caer en el mismo óptimo local, es decir, a la misma solución.

En la Figura 2 se presenta el método propuesto en pseudocódigo.

```

BEGIN
 $\overline{cx}_j = \overline{cy}_j = 0 \quad j = 1, \dots, 20$ 
 $best = +\infty$ 
REPEAT

$$z(x) = \sum_{j=1}^{20} \sum_{f \in F} \sum_{l \in L_f} x_{l,g_f(j)}^f \sqrt{(u_l - \overline{cx}_j)^2 + (v_l - \overline{cy}_j)^2}$$

SOLVE ILP usando  $z(x)$  como función objetivo

$$\overline{cx}_j = \frac{\sum_{f \in F} \sum_{l \in L_f} u_l * x_{l,g_f(j)}^f}{\sum_{f \in F} \sum_{l \in L_f} x_{l,g_f(j)}^f} \quad j = 1, \dots, 20$$


$$\overline{cy}_j = \frac{\sum_{f \in F} \sum_{l \in L_f} v_l * x_{l,g_f(j)}^f}{\sum_{f \in F} \sum_{l \in L_f} x_{l,g_f(j)}^f} \quad j = 1, \dots, 20$$

IF  $z(x) < best$ 
 $best = z(x)$ 

$$\overline{x}_{l,g_f(j)}^f = x_{l,g_f(j)}^f \quad j = 1, \dots, 20 \quad f \in F \quad l \in L_f$$

IF  $x$  es óptimo local
 $\overline{cx}_j = random \quad j = 1, \dots, 20$ 
 $\overline{cy}_j = random \quad j = 1, \dots, 20$ 
UNTIL condición de termino
END

```

Figura 2: Algoritmo heurístico propuesto en pseudocódigo.

- **Diseño de experimento para evaluar el método de solución propuesto.**

Estructuras del problema consideradas.

Los valores fijados a determinados parámetros como son: tamaño de las instancias, estructura de las instancias, número de *clusters* a obtener, tienen en cuenta no solo la personalización al problema inmediato a resolver por la empresa, el agrupamiento de clientes de la empresa en una ciudad, sino la solución que en un futuro cercano debe darse en otras regiones del país, donde estos parámetros pudieran cambiar significativamente. Por ejemplo, en la actualidad en una ciudad grande el número de clientes oscila en el orden de 1000, el número de zonas de distribución es 10, pero se valoraron situaciones diferentes en otras provincias o regiones del país.

Los algoritmos fueron evaluados en un conjunto de instancias tomadas de la literatura relacionada con la experimentación que se aborda (Homberger (1998)), para problemas afines al presentado por la empresa, así como en instancias generadas con las características que presenta el problema abordado.

Los elementos que se tuvieron en cuenta de las instancias, para encontrar una posible correlación con el desempeño de los algoritmos desarrollados, son:

1. Tamaño de las instancias, considerando como tal el número de datos (clientes) a agrupar.
2. Estructura de las instancias, es decir, las características de la distribución geográfica de los clientes en las zonas a particionar.

En relación a la estructura de las instancias se tuvieron en cuenta las tres clases de instancias según se mostró en el epígrafe de construcción de *clusters*.

Las instancias varían el rango del tamaño (número total de clientes) entre 100 y 1000. Se tomaron instancias de tamaño 100, 200, 300, 400, 600, 800 y 1000 basadas en las instancias propuestas en (Homberger (1998)), y que consideran además las clases descritas previamente.

La distribución de frecuencias sobre el total de clientes en cada una de sus instancias (BS,S,Q,M) considera la situación actual de 10 regiones como se muestra en la Tabla 1, es decir, distribuir las frecuencias (bisemanal, semanal, quincenal y mensual) con la misma probabilidad de las 10 zonas de los datos reales que se tienen.

Zona	BS	S	Q	M
z1	0,00%	56,61%	30,17%	13,22%
z2	0,92%	52,53%	36,87%	9,68%
z3	0,67%	45,30%	39,26%	14,77%
z4	1,48%	35,93%	55,19%	7,41%
z5	6,49%	45,41%	40,54%	7,57%
z6	1,88%	54,46%	40,85%	2,82%
z7	0,93%	35,81%	47,44%	15,81%
z8	0,00%	27,57%	67,65%	4,78%
z9	0,00%	54,77%	40,66%	4,56%
z10	0,00%	59,68%	8,06%	32,26%

Tabla 1: Distribución de las frecuencias de entrega.

Condiciones de término de las soluciones heurísticas obtenidas:

- a. detenerse al encontrar el primer óptimo local (*iter1*).
- b. detenerse al encontrar el segundo óptimo local (*iter2*).
- c. detenerse después de encontrar 1 óptimo local consecutivo sin mejorar el óptimo global (*nome1*).
- d. detenerse después de encontrar 5 óptimos locales consecutivos sin mejorar el óptimo global (*nome5*).
- e. detenerse después de encontrar 100 óptimos locales (*iter100*) explorando una gran cantidad de óptimos locales diferentes posibles, usando un tiempo de cálculo elevado. Esta condición de término es usada como la mejor solución conocida para poder comparar la calidad de los otros criterios.

La implementación fue desarrollada en un computador con un procesador Intel Pentium 4 de 3.40GHz usando como sistema operativo Ubuntu Linux 11.10. Para resolver los modelos ILP de forma exacta se usó el solucionador IBM ILOG CPLEX 12.4.

Resultados experimentales.

Se realizaron extensas pruebas computacionales para evaluar el desempeño de la heurística propuesta en términos de la calidad de solución (%gap) respecto a la mejor solución conocida como del tiempo de cómputo requerido para alcanzarla.

Se generaron 350 instancias que fueron generadas a partir de cinco grupos de coordenadas para los clientes con diversa distribución, ó clase, que fueron presentados por (Hombberger (1998)), considerando 100, 200, 300, 400, 600, 800 y 1000 clientes para cada uno de estos grupos. Finalmente a cada cliente se le asoció una frecuencia de visita de acuerdo a las probabilidades de cada una de las diez zonas de distribución que se muestran en la Tabla 1.

En las Tablas 2, 3 y 4 se pueden observar las cinco condiciones de término evaluadas, presentando para cada una la diferencia entre su solución y la mejor solución encontrada (%gap), el tiempo de cómputo utilizado expresado en segundos (CPU Time) y el número de óptimos locales visitados (Iter.).

De acuerdo a la Tabla 2, si agrupamos las soluciones de acuerdo al número de clientes podemos observar que a medida que aumenta el número de iteraciones, el tiempo de cómputo requerido aumenta consistentemente. Así mismo, la calidad de la solución mejora en la medida que aumenta el número de iteraciones, esto se refleja con la disminución del %gap.

El aumento del tiempo de cómputo respecto al número de clientes es debido a que el modelo ILP que se soluciona de forma exacta en cada iteración aumenta su número de variables binarias según aumenta el número de clientes.

Número de clientes	iter1			iter2			nome1			nome5			iter100		
	%GAP	CPU Time	Iter.	%GAP	CPU Time	Iter.	%GAP	CPU Time	Iter.	%GAP	CPU Time	Iter.	%GAP	CPU Time	Iter.
100	7,76%	0,71	1,00	2,89%	1,53	2,00	2,01%	3,09	2,92	0,75%	8,88	9,70	0,00%	96,98	100,00
200	4,45%	2,04	1,00	2,47%	4,34	2,00	1,96%	7,97	2,60	0,67%	23,45	8,72	0,00%	306,92	100,00
300	4,23%	5,20	1,00	2,45%	10,28	2,00	2,16%	16,12	2,60	0,70%	50,98	9,26	0,00%	643,02	100,00
400	3,54%	7,00	1,00	1,90%	14,18	2,00	1,73%	23,54	2,66	0,64%	73,43	9,26	0,00%	966,40	100,00
600	2,52%	12,47	1,00	1,96%	27,28	2,00	1,62%	40,37	2,62	0,35%	166,88	10,30	0,00%	1.871,33	100,00
800	3,00%	20,11	1,00	1,39%	44,37	2,00	1,16%	74,12	2,72	0,20%	269,49	10,30	0,00%	3.081,42	100,00
1000	2,22%	25,61	1,00	1,32%	57,06	2,00	1,10%	93,74	2,60	0,32%	358,43	10,16	0,00%	4.046,23	100,00
Promedio	3,96%	10,45	1,00	2,06%	22,72	2,00	1,68%	36,99	2,67	0,52%	135,93	9,67	0,00%	1.573,19	100,00

Tabla 2: Resumen de resultados con promedios por número de clientes.

Tipos	iter1			iter2			nome1			nome5			iter100		
	%GAP	CPU Time	Iter.	%GAP	CPU Time	Iter.	%GAP	CPU Time	Iter.	%GAP	CPU Time	Iter.	%GAP	CPU Time	Iter.
c1	4,92%	9,82	1,00	2,25%	21,69	2,00	1,56%	34,03	2,86	0,53%	122,88	9,17	0,00%	1.505,73	100,00
c2	5,05%	10,74	1,00	2,32%	22,53	2,00	2,05%	36,36	2,69	0,50%	135,69	10,41	0,00%	1.608,46	100,00
r1	2,29%	10,88	1,00	1,55%	23,07	2,00	1,36%	36,69	2,41	0,49%	154,47	9,89	0,00%	1.691,89	100,00
r2	3,33%	10,02	1,00	1,86%	23,67	2,00	1,68%	39,61	2,73	0,57%	142,49	9,30	0,00%	1.552,44	100,00
rc1	4,21%	10,78	1,00	2,30%	22,64	2,00	1,74%	38,27	2,69	0,50%	124,15	9,59	0,00%	1.507,41	100,00
Promedio	3,96%	10,45	1,00	2,06%	22,72	2,00	1,68%	36,99	2,67	0,52%	135,93	9,67	0,00%	1.573,19	100,00

Tabla 3: Resumen de resultados con promedios por tipo de instancia.

Zonas	iter1			iter2			nome1			nome5			iter100		
	%GAP	CPU Time	Iter.	%GAP	CPU Time	Iter.	%GAP	CPU Time	Iter.	%GAP	CPU Time	Iter.	%GAP	CPU Time	Iter.
z1	1,28%	8,54	1,00	0,68%	20,18	2,00	0,50%	37,46	2,71	0,20%	104,17	8,89	0,00%	1.270,79	100,00
z2	3,67%	10,86	1,00	1,82%	23,35	2,00	1,52%	34,76	2,74	0,53%	111,46	8,49	0,00%	1.482,21	100,00
z3	4,59%	9,75	1,00	1,92%	20,06	2,00	1,78%	32,05	2,54	0,58%	123,18	9,97	0,00%	1.301,38	100,00
z4	5,40%	9,10	1,00	2,79%	21,73	2,00	1,89%	39,52	2,83	0,81%	173,40	9,97	0,00%	1.693,38	100,00
z5	6,03%	11,49	1,00	3,83%	24,40	2,00	3,15%	35,97	2,54	0,88%	150,74	9,71	0,00%	1.796,03	100,00
z6	3,78%	7,84	1,00	2,63%	18,72	2,00	2,37%	30,96	2,66	0,58%	119,58	9,23	0,00%	1.440,43	100,00
z7	3,97%	10,19	1,00	2,57%	21,83	2,00	2,53%	38,34	2,51	0,88%	135,02	10,11	0,00%	1.570,63	100,00
z8	5,46%	17,57	1,00	2,30%	36,46	2,00	1,48%	56,90	2,89	0,40%	212,99	10,46	0,00%	2.490,33	100,00
z9	3,10%	10,93	1,00	1,01%	23,31	2,00	0,81%	33,94	2,51	0,18%	125,11	9,77	0,00%	1.494,01	100,00
z10	2,32%	8,20	1,00	1,01%	17,16	2,00	0,74%	30,01	2,80	0,17%	103,68	10,11	0,00%	1.192,67	100,00
Promedio	3,96%	10,45	1,00	2,06%	22,72	2,00	1,68%	36,99	2,67	0,52%	135,93	9,67	0,00%	1.573,19	100,00

Tabla 4: Resumen de resultados con promedios por zonas.

En la Tabla 3 podemos observar que no se presentan diferencias significativas para la calidad de la solución si los clientes están distribuidos de acuerdo a las diferentes clases c , r y rc .

En la Tabla 4 podemos observar que las diferentes zonas, cada una con una probabilidad de frecuencia de visita diferente, afecta la calidad de solución de nuestro algoritmo.

De las Tablas 2, 3 y 4 puede observarse que entre las condiciones de término $nome5$ y $iter100$ no existe una diferencia importante en la calidad de solución, pero el tiempo de cómputo es un orden de magnitud menor con $nome5$; por ello, para encontrar un balance entre el tiempo de cómputo utilizado y la calidad de la solución, es recomendable el uso de la condición de término $nome5$.

La Figura 3 muestra gráficamente cómo evoluciona la función objetivo. En cada iteración se observa una mejora del valor de la función objetivo hasta un punto en donde queda atrapado en un óptimo local; cuando esto sucede se ejecuta el procedimiento de diversificación para escapar y explorar otras soluciones de la región factible. Este procedimiento termina de acuerdo a la condición de término utilizada.

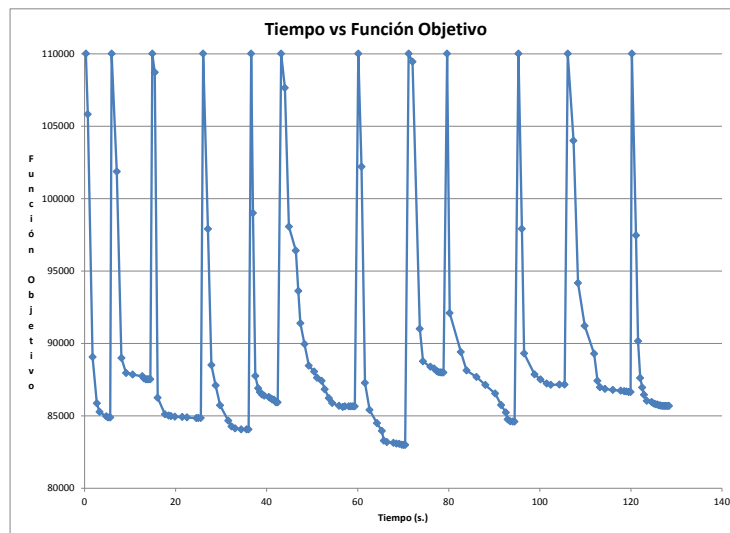


Figura 3: Valor de la función objetivo respecto al número de iteraciones.

La Figura 4 muestra un ejemplo pequeño con clientes de frecuencia semanal. Se observa en cada una de las cinco trayectorias, que corresponden a los centroides de los cinco días de la semana, su evolución a medida que avanza el algoritmo. Al comienzo todos los centroides se encuentran agrupados en un mismo punto al centro y se van desplazando hacia los *clusters* que forman los clientes, en la última iteración los centroides se ubican exactamente en el centro del *cluster* que corresponde a la solución óptima para este ejemplo.

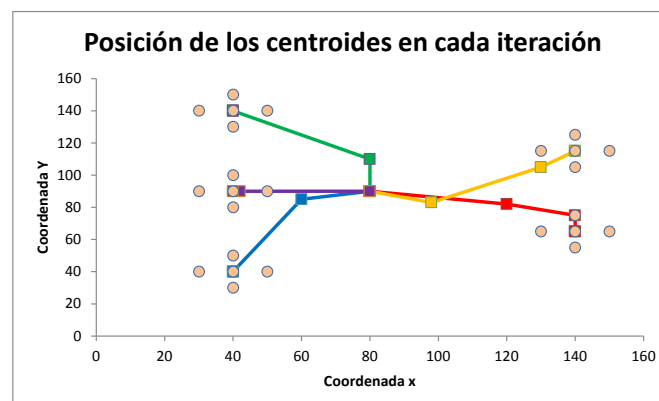


Figura 4: Posición de los centroides en las diferentes iteraciones del algoritmo.

Conclusiones y trabajo futuro.

El trabajo desarrollado plantea una propuesta de solución a un problema de la vida real logístico altamente complejo, no abordado como tal en la literatura previamente.

Los resultados obtenidos de agrupamiento de clientes y de planificación de visitas mensual a los mismos han satisfecho a los directivos de la empresa, en particular a la gerencia de ventas. El diseño de experimentos fue realizado con tipos de instancias que solicitó la empresa según situaciones actuales y otras que deben abordar en un futuro cercano.

Como trabajo inmediato se encuentra abordar la fase de ruteo, e integrarlo en un sistema de apoyo a la toma de decisiones, sustentado por un sistema de información geográfica.

Referencias bibliográficas.

- Averbakh, I. and Berman, O.** (1999), A simple heuristic for machine flow-shop and its applications in routing-scheduling problems. *Operations Research* 47, 165-170.
- Bar-Ilan, J., Kortsarz, G. and Peleg, D.** (1993), How to allocate network centers, *J. Algorithms*, 15(3), 385--415
- Caprara, A., Monaci, M. and Toth, P.** (2001), A global method for crew planning in railway applications. In: Daduna, J., Voss, S. (eds.) *Computer-Aided Transit Scheduling. Lecture Notes in Economics and Mathematical Systems*, 505, 17–36. Springer, New York.
- Chakroborty, P.** (2003), Genetic algorithms for optimal urban transit network design. *Comput. Aided Civil Infrastruct. Eng.* 18(3), 184–200.
- Cortés, M.E. and García, R.A.** (2007), Solución al problema de asignación-distribución. *Mosaicos Matemáticos Memorias de la XVII Semana Regional de Investigación y Docencia en Matemáticas*, Universidad de Sonora México, 20, 51-57.
- Deng, Y. and Bard J.F.** (2010), A reactive GRASP with path relinking for capacitated clustering. Springer Science + Business Media.
- Escalera, A.** (2011), Sistema soporte a la decisión para el agrupamiento en zonas de distribución de los clientes de BRASCUBA S.A., Tesis de maestría en Ciencias en Informática Empresarial, La Habana, Cuba, Mayo 2011.
- Freling, R., Huisman, D. and Wagelmans, A.P.M.** (2003), Models and algorithms for integration of vehicle and crew scheduling. *J. Sched.* 6(1), 63–85.
- González, G. and González, F.** (2007), Metaheurísticas aplicadas al ruteo de vehículos. Un caso de estudio. Parte 3: Genetic clustering and Tabu routing. 27, 106-113.
- Homberger, J.** (1998), Extended SOLOMON's VRPTW instances, <http://www.fernuni-hagen.de/WINF/touren/inhalte/probinst.htm>, visitado el 10/ene/2012.
- Jayaraman, V.** (1998), Transportation, facility location and inventory issues in distribution network design, *International Journal of Operations & Production Management*, 18(5), 471-494.
- Korsvik, J.E. and Fagerholt, K.** (2010), A tabu search heuristic for ship routing and scheduling with flexible cargo quantities, *Journal of Heuristics* 16(2), 117-137.
- Letchford, A. and Burer, S.** (2012), Non-convex mixed-integer nonlinear programming: a survey, *Surveys in Operations Research and Management Science*, In Press.
- Nagy, G. and Salhi S.** (2007), Location-routing: Issues, models and methods, *European Journal of Operational Research* 177(2), 649-672.
- Pacheco, J. and Beltrán M.** (2001), Análisis de nuevos métodos de clasificación. Un ejemplo ilustrativo de su uso en la agrupación de los municipios de Castilla y León, *Estadística Española* Vol. 43(148), 209-224.
- Pentico, D.W.** (2007), Assignment problems: A golden anniversary survey, 176(2), 774-793.
- Taillard, E.** (2003), Heuristic Methods for Large Centroid Clustering Problems, *Journal of Heuristics* 9(1), 51-73.
- Zhao, B. Y., J. Kubiawicz, and Joseph, A.D.** (2001), Tapestry: An infrastructure for fault-tolerant wide-area location and routing, *Computer* 74 (11-20), 46.