

Heurísticos de Búsqueda

Jon Goenetxea

23 de febrero de 2015

Resumen

En este trabajo se presenta una comparativa entre dos diferentes algoritmos de búsqueda heurística. El primero, denominado *Hill Climbing*, es un algoritmo basado en búsqueda local, que se ha implementado utilizando múltiples inicializaciones (*multi-start*) para tratar de evitar los mínimos locales. Como segundo algoritmo para la comparación se ha escogido un algoritmo genético, que entra dentro de los algoritmos tipificados como poblacionales. En el apartado experimental, se han analizado las posibles configuraciones para la búsqueda local, y se ha hecho una comparación utilizando un test de hipótesis de Wilcoxon.

1. Problema de asignación cuadrática

El problema de la asignación cuadrática, que se denota por sus siglas en inglés *QAP* (*Quadratic assignment problem*), fue planteado como un modelo matemático para un conjunto de actividades económicas indivisibles. Posteriormente fue demostrado que QAP pertenece a los problemas no polinomiales duros, y es aplicable a un sin número de situaciones.

La primera vez que este problema fue planteado fue en 1957 por Koopmans y Beckmann. En 1976, Sahni y Gonzalez demostraron que era un problema NP duro.

Este tipo de problemas consisten en escoger para N localizaciones la distribución de N instalaciones que minimice el coste de instalación. Para la toma de decisiones hay que tener en cuenta dos factores principales: 1) el coste de moverse entre localizaciones (distancias, coste de transporte, penalizaciones, etc.), y 2) el flujo de material entre instalaciones.

Para definir estos datos, se definen dos matrices (la matriz de distancias y la matriz de flujos de material) que modelan el comportamiento del sistema de instalaciones. La matriz de distancias se define como $D = [d_{ij}]$, donde d_{ij} representa el coste de transporte de la localización i a la localización j . La matriz de flujos se define como $F = [f_{kl}]$, que representa el flujo de material entre la instalación k a la instalación l .

Cada configuración (una posible ubicación para todas las instalaciones) se representa en forma de permutación (ver ecuación 1).

$$\sigma = (\sigma_1 \sigma_2 \sigma_3 \dots \sigma_n) \quad (1)$$

La función de coste a minimizar por lo tanto, tendrá en cuenta el coste de transporte dependiendo del flujo entre las instalaciones y la distancia entre las

localizaciones. Esta función se puede modelar como muestra la ecuación 2.

$$f(\sigma) = \sum_{i=1}^n \sum_{j=1}^n d_{\sigma(i)\sigma(j)} f_{ij} \quad (2)$$

2. Propuestas de solución basadas en búsqueda local

Como se ha presentado anteriormente, el algoritmo implementado para la búsqueda local es el *Hill Climbing*. Es un algoritmo de búsqueda local sencillo. Se basa en analizar los resultados vecinos de una muestra, en busca de un resultado mejor que el actual. Esto se hace iterativamente, por lo que el caer en un óptimo local es muy probable. Para minimizar el caer en óptimos locales, se hacen varias reinicializaciones (conocido como *multi start*) utilizando el mejor de los resultados como respuesta final.

El pseudo-código para este algoritmo sería el siguiente:

Data: cantidad de 'start'-s, función de coste

Result: valor de error mínimo encontrado

```

for cada 'start' do
    generar permutación inicial con función aleatoria;
    while se ha encontrado mejor permutación do
        calcular vecindario;
        for cada vecino do
            calcular valor del vecino;
            if mejora then
                guardar nuevo vecino como actual;
                se ha encontrado mejor permutación;
            end
        end
    end
end
devolver valor de mejor permutación encontrada

```

Algorithm 1: Pseudocódigo de *Hill Climbing*

Este algoritmo tiene pocos parámetros para configurar, lo que simplifica su configuración. El único parámetro que es necesario definir es la cantidad de inicializaciones a hacer. No obstante, es necesario definir el método de calcular los vecindarios, y hay que escoger uno que se adecue al problema.

En este aspecto se han analizado dos sistemas de vecinos diferentes: *Swap* y *2-opt*. El sistema de vecinos *Swap* consiste en generar vecinos cambiando un elemento de la lista con otro. Si intercambiamos un elemento al azar con el resto, generamos un vecindario con n vecinos. El sistema de vecinos *2-opt* consiste en intercambiar dos elementos en la lista de elementos. Si se tienen en cuenta todas las posibles permutaciones, se consigue un vecindario de k elementos, donde k se calcula con la ecuación 3.

Los experimentos realizados muestran que

$$k = \frac{n * (n - 1)}{2} - n \quad (3)$$

En los experimentos, se ha comprobado que el algoritmo *2-Opt* es más efectivo a la hora de acercarse a la solución óptima. No obstante, el tiempo de computo también es significativamente mayor que el del método *Swap*. A pesar de esta desventaja, se ha optado por utilizar el método *2-Opt*.

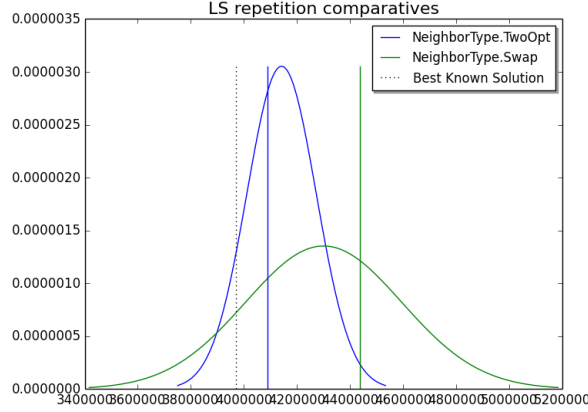


Figura 1: Comparativa entre 2-Opt y Swap, con 30 repeticiones y cogiendo los mejores resultados en cada una.

3. Propuestas de solución basadas en algoritmos poblacionales

El algoritmo poblacional implementado es un algoritmo genético. Este algoritmo se basa en las teorías clásicas de la evolución darwiniana como base para generar una serie de iteraciones que generan una población mejor que la generación anterior.

Un algoritmo genético consta de cuatro fases principales: 1) evaluación (evaluation), 2) selección (selection), 3) cruce (crossover) y 4) mutación (mutation). Cada iteración genera una nueva generación después de pasar por todos estos estados. Se puede poner un criterio de terminación, con el que podemos decidir si la generación actual es ya lo suficientemente buena para el resultado, o si se han utilizado los recursos asignados, por ejemplo. La figura 2 muestra gráficamente la secuencia de estados de una iteración.

3.1. Fase de Evaluación

En la fase de evaluación se evalúan todos los individuos de la población actual. Esto se hace utilizando la función de coste que se ha expuesto en el apartado 1.

Por lo tanto, cada individuo tendrá asociado el valor de su resultado, dando un indicador de la "calidad" de dicho individuo.

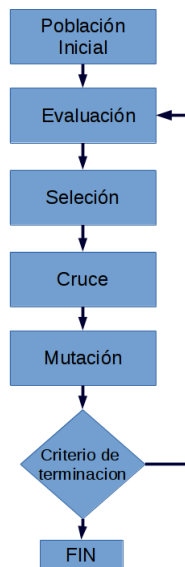


Figura 2: Secuencia de una iteración de un algoritmo genético.

3.2. Fase de Selección

En la fase de selección se escogen los individuos de la población con mejores resultados. Para ello se pueden utilizar diferentes opciones. Entre ellas se ha escogido la *selección por torneo*.

Esta selección se realiza mediante una comparación entre un subconjunto de individuos de tamaño concreto (definido como parámetro), escogidos al azar dentro de la población.

Este tipo de selección se efectúa de manera muy rápida dado que no es necesario evaluar la totalidad de la población. No obstante, hay que definir un nuevo parámetro por lo que añade complejidad a la configuración.

3.3. Fase de Cruce

En esta fase, se unen y combinan los diferentes individuos en el subconjunto de *mejores* individuos. Es necesario encontrar una función de cruce que genere nuevos individuos aprovechando las cualidades de los originales, y tratando de transmitirlos a las nuevas generaciones. Para esta fase, se han barajado varias opciones.

La primera de ellas, combina dos individuos por mitades. Es decir, coge la primera mitad de un individuo A, y la segunda mitad de un individuo B. Los nuevos individuos (denominados A' y B') tendrán la primera parte de A y la segunda de B respectivamente. Para rellenar la mitad restante, lo que se hace es colocar los valores que faltan por incluir con el mismo orden que aparece en el individuo original. Por lo tanto, A' tendrá la primera mitad de A, y los elementos restante en el mismo orden que aparecen en B, y B' tendrá la segunda mitad de B y los demás valores en el mismo orden que aparecen en A.

Como segunda opción, se ha planteado una función en la que se intercambia

Parámetro	Valor
Tamaño población	150
Número de generaciones	15
Factor Mutación	0.05
Tamaño subconjunto torneo	3
Factor de cruce	0.5

Cuadro 1: Lista de parámetros de configuración para el algoritmo genético.

un elemento al azar entre los dos individuos. Es decir, se escoge un índice al azar (por ejemplo 2) y se coloca el valor en 2 de A en la posición 2 de B, y viceversa. Esta acción por si sola es probable que duplique valores en el nuevo individuo. Para evitar esto, el antiguo valor en la posición 2 se coloca en el lugar en el que está el valor que se va a insertar. La figura 3 representa gráficamente el proceso de cruce descrito.

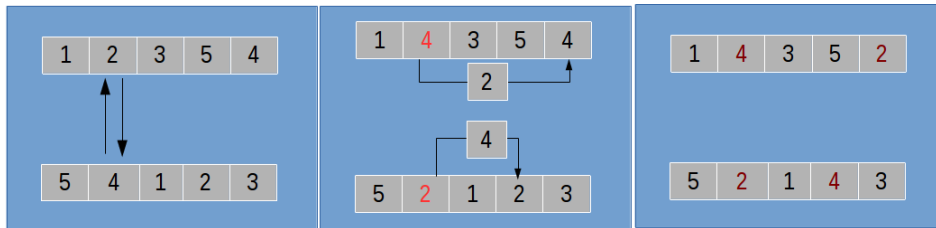


Figura 3: La figura muestra de forma gráfica el proceso de cruce escogido para el algoritmo genético. En el primer paso (izquierda), se seleccionan los elementos con un índice escogido al azar. En el segundo (centro), se intercambia el elemento entre los dos individuos, y se detecta la posición del número que va a ser sustituido. El tercer paso (derecha), se pueden ver los nuevos individuos, y como no tienen repeticiones.

3.4. Fase de Mutación

Para aumentar la variedad de posibilidades, se añade un factor de mutación. Este factor hace que una parte de la población mute (tenga un cambio adicional) en cada iteración. Con esto conseguimos configuraciones que no estuviesen en la generación anterior, y que la fase de cruce no puede generar.

La función de mutación escogida es la de intercambiar 2 elementos del individuo. Para ello, se escogen dos índices al azar, y se intercambian los elementos.

Esta fase añade un elemento a configurar, que es la probabilidad de mutación de un individuo.

3.5. Configuración final

A diferencia de el algoritmo de *Hill Climbing*, este tipo de algoritmos tiene más parámetros que definir. La tabla 1 lista los parámetros escogidos para este algoritmo.

Se ha definido el número de generaciones para que sea el mismo que la cantidad de iteraciones que se realizan en el algoritmo de búsqueda local. En este sentido, tanto el algoritmo de búsqueda local realizará n iteraciones, y el algoritmo genético creará a su vez n generaciones para lograr el resultado final.

En el caso de la búsqueda local, se escoge el mejor valor de todas las iteraciones. En el caso de los algoritmos genéticos, la última generación contiene por definición el mejor de los valores calculados. De esta manera, igualamos la cantidad de ciclos de cada uno, y podemos comparar los resultados y los tiempos de ejecución de ambos algoritmos.

4. Experimentación

5. Conclusiones