

UNIVERSIDAD DE GRANADA
E.T.S.I. INFORMÁTICA Y TELECOMUNICACIÓN



**UNIVERSIDAD
DE GRANADA**

Departamento de Ciencias de la
Computación e Inteligencia Artificial

Metaheurísticas

Guión de Prácticas

Práctica 1.a:
Técnicas de Búsqueda Local y Algoritmos Greedy
para el Problema de la Máxima Diversidad

Curso 2020-2021

Tercer Curso del Grado en Ingeniería Informática

Práctica 1.a

Técnicas de Búsqueda Local y Algoritmos Greedy para el Problema de la Máxima Diversidad

1. Objetivos

El objetivo de esta práctica es estudiar el funcionamiento de las *Técnicas de Búsqueda Local* y de los *Algoritmos Greedy* en la resolución del problema de la máxima diversidad (MDP) descrito en las transparencias del Seminario 2. Para ello, se requerirá que el estudiante adapte los siguientes algoritmos a dicho problema:

- Algoritmo Greedy básico.
- Algoritmo de Búsqueda Local (BL).

El estudiante deberá comparar los resultados obtenidos con las estimaciones existentes para el valor de los óptimos de una serie de casos del problema.

La práctica se evalúa sobre un total de **2 puntos**, distribuidos de la siguiente forma:

- BL (**1.5 puntos**).
- Greedy (**0.5 puntos**).

La fecha límite de entrega será el **domingo 11 de abril de 2021** antes de las 23:55 horas. La entrega de la práctica se realizará por internet a través del espacio de la asignatura en PRADO.

2. Trabajo a Realizar

El estudiante podrá desarrollar los algoritmos de la práctica siguiendo la modalidad que desee: trabajando con cualquier *framework* de metaheurísticas existente, implementándolos a partir del código C proporcionado en la web de la asignatura o considerando cualquier código disponible en Internet.

Los métodos desarrollados serán ejecutados sobre una serie de casos del problema. Se realizará un estudio comparativo de los resultados obtenidos y se analizará el comportamiento de cada algoritmo en base a dichos resultados. **Este análisis influirá decisivamente en la calificación final de la práctica.**

En las secciones siguientes se describen los aspectos relacionados con cada algoritmo a desarrollar y las tablas de resultados a obtener.

3. Problema y Casos Considerados

3.1. Introducción al Problema de la Máxima Diversidad

El problema de la máxima diversidad (en inglés, *maximum diversity problem*, MDP) es un problema de optimización combinatoria consistente en seleccionar un subconjunto M de m elementos ($|M|=m$) de un conjunto inicial N de n elementos (con $n>m$) de forma que se maximice la diversidad entre los elementos escogidos. El MDP se puede formular como:

$$\begin{aligned} \text{Maximizar } z_{MS}(x) &= \sum_{i=1}^{n-1} \sum_{j=i+1}^n d_{ij} x_i x_j \\ \text{Sujeto a } \sum_{i=1}^n x_i &= m \\ x_i &= \{0, 1\}, \quad i = 1, \dots, n. \end{aligned}$$

donde:

- x es una solución al problema que consiste en un vector binario que indica los m elementos seleccionados
- d_{ij} es la distancia existente entre los elementos i y j .

3.2. Casos del MDP Considerados

Se utilizarán **30 casos** seleccionados de varios de los conjuntos de instancias disponibles en la MDPLIB (<https://grafo.etsii.urjc.es/opticom/mdp/>), todas pertenecientes al grupo **MDG** con distancias aleatorias:

- 10 del grupo **MDG-a** con distancias enteras en $\{0,10\}$, $n=500$ y $m=50$ (*MDG-a_1_n500_m50* a *MDG-a_10_n500_m50*),
- 10 del grupo **MDG-b** con distancias reales en $[0,1000]$, $n=2000$ y $m=20$ (*MDG-b_21_n2000_m200* a *MDG-b_30_n2000_m200*); y
- 10 del grupo **MDG-c** con distancias enteras en $\{0,1000\}$, $n=3000$ y $m=\{300, 400, 500, 600\}$ (*MDG-c_1_n3000_m300* a *MDG-c_20_n3000_m600*).

El fichero *Tablas_MDP_2020-21.xls*, disponible en el espacio de PRADO y la web de la asignatura, incluye las características (nombre, tamaño y coste de la mejor solución conocida) de las instancias seleccionadas. Aunque pueden ser descargados de la MDPLib, los ficheros de los 30 casos se han incluido también en las dos

localizaciones del espacio de la asignatura (PRADO y web externa) para facilitar el trabajo al estudiantado.

El formato de los ficheros es el siguiente:

- Una primera línea donde se indica el número de elementos n y el número de elementos a seleccionar m del problema.
- $n \cdot (n-1)/2$ líneas con el formato $i \ j \ d_{ij}$ ($i, j \in \{0, \dots, n-1\}$) que recogen el contenido de la matriz de distancias entre los elementos. Se almacena sólo la diagonal superior.

Ejemplo (*GKD-c_1_n500_m50*):

```
500 50
0 1 11.17945
0 2 12.18565
...
496 497 14.48240
496 498 11.37189
496 499 13.94453
497 498 15.47191
497 499 17.05433
498 499 10.37931
```

4. Algoritmos

4.1. Algoritmo *Greedy*-MDP

El algoritmo *greedy* del MDP se basa en la heurística de ir seleccionando los elementos más lejanos a los previamente seleccionados. Para ello, se parte del elemento más alejado del resto en el conjunto completo de elementos. En los $m-1$ pasos siguientes se va escogiendo el elemento más lejano a los elementos seleccionados hasta el momento.

4.2. Búsqueda Local

Algoritmo

Como algoritmo de BL para el MDP consideraremos el esquema del primer mejor, tal y como está descrito en las transparencias del Seminario 2. La representación será en forma de un conjunto de elementos seleccionados. Se empleará el movimiento de intercambio $Int(Sel, i, j)$ que intercambia el elemento seleccionado i por uno no seleccionado j en la solución actual Sel . **Será obligatorio emplear la factorización** de la función objetivo en todos los casos. Una vez realizado el movimiento, se actualiza la solución actual y los valores de contribución de los elementos seleccionados al coste de dicha solución, y se comienza a explorar el nuevo entorno. La exploración del entorno se realizará en orden ascendente de contribución de los elementos seleccionados al coste de la solución actual.

En cada ejecución de la BL, se partirá de una solución inicial aleatoria y se detendrá la ejecución **cuando no se encuentre mejora en todo el entorno o cuando se hayan realizado 100000 evaluaciones de la función objetivo**, es decir, en cuanto se cumpla alguna de las dos condiciones. Se realizará una única ejecución sobre cada caso del problema.

4.3. Determinación de la Calidad de un Algoritmo

El modo habitual de determinar la calidad de un algoritmo de resolución aproximada de problemas de optimización es ejecutarlo sobre un conjunto determinado de instancias y comparar los resultados obtenidos con los mejores valores conocidos para dichas instancias.

Además, los algoritmos pueden tener diversos parámetros o pueden emplear diversas estrategias. Para determinar qué valor es el más adecuado para un parámetro o saber la estrategia más efectiva los algoritmos también se comparan entre sí.

La comparación de los algoritmos se lleva a cabo fundamentalmente usando dos criterios, la *calidad* de las soluciones obtenidas y el *tiempo de ejecución* empleado para conseguirlas. Además, es posible que los algoritmos no se comporten de la misma forma si se ejecutan sobre un conjunto de instancias u otro.

Por otro lado, a diferencia de los algoritmos determinísticos, los algoritmos probabilísticos se caracterizan por la toma de decisiones aleatorias a lo largo de su ejecución. Este hecho implica que un mismo algoritmo probabilístico aplicado al mismo caso de un problema pueda comportarse de forma diferente y por tanto proporcionar resultados distintos en cada ejecución.

Cuando se analiza el comportamiento de un algoritmo probabilístico en un caso de un problema, se desearía que el resultado obtenido no estuviera sesgado por una secuencia aleatoria concreta que pueda influir positiva o negativamente en las decisiones tomadas durante su ejecución. Por tanto, resulta necesario efectuar varias ejecuciones con distintas secuencias probabilísticas y calcular el resultado medio y la desviación típica de todas las ejecuciones para representar con mayor fidelidad su comportamiento.

Dada la influencia de la aleatoriedad en el proceso, es recomendable disponer de un generador de secuencia pseudoaleatoria de buena calidad con el que, dado un valor semilla de inicialización, se obtengan números en una secuencia lo suficientemente grande (es decir, que no se repitan los números en un margen razonable) como para considerarse aleatoria. En el espacio de PRADO y la web de la asignatura se puede encontrar una implementación en lenguaje C de un generador aleatorio de buena calidad (*random.h*).

Como norma general, el proceso a seguir consiste en realizar un número de ejecuciones diferentes de cada algoritmo probabilístico considerado para cada caso del problema. Es necesario asegurarse de que se realizan diferentes secuencias aleatorias en dichas ejecuciones. Así, el valor de la semilla que determina la inicialización de cada secuencia deberá ser distinto en cada ejecución y estas semillas deben mantenerse en los distintos algoritmos (es decir, la semilla para la primera ejecución de todos los

algoritmos debe ser la misma, la de la segunda también debe ser la misma y distinta de la anterior, etc.). Para mostrar los resultados obtenidos con cada algoritmo en el que se hayan realizado varias ejecuciones, se deben construir tablas que recojan los valores correspondientes a estadísticos como el **mejor** y **peor** resultado para cada caso del problema, así como la **media** y la **desviación típica** de todas las ejecuciones. También se pueden emplear descripciones más representativas como los boxplots, que proporcionan información de todas las ejecuciones realizadas mostrando mínimo, máximo, mediana y primer y tercer cuartil de forma gráfica. Finalmente, se construirán unas tablas globales con los resultados agregados que mostrarán la calidad del algoritmo en la resolución del problema desde un punto de vista general.

Alternativamente, **en el caso de que se considere un número alto de casos del problema, se puede confiar en una única ejecución del algoritmo. Esta condición se cumple en las prácticas que realizaremos con el MDP.** Aun así, **será necesario inicializar la semilla del generador aleatorio para poder repetir el experimento** y obtener los mismos resultados si fuera necesario (en caso contrario, los resultados podrían variar en cada ejecución del mismo algoritmo sobre el mismo caso del problema).

Para facilitar la comparación de algoritmos en las prácticas del MDP se considerarán dos estadísticos distintos denominados *Desv* y *Tiempo*:

- *Desv* es la media de las desviaciones, en porcentaje, del valor obtenido en cada ejecución respecto al mejor valor conocido para cada caso del problema. En el caso de problemas de maximización como el MDP: quitar el -g cuando se vayan a tomar los valores.

$$\frac{Valor_i - ValorAlgoritmo_i}{MejorValor_i}$$

no hay que medir el tiempo de ejecución del algoritmo, no hay que incluir la solución para una instancia, sino que se determina lo cerca que se está de obtener la mejor solución conocida.

- *Tiempo* en segundos si se quiere como la media de los tiempos de ejecución de los algoritmos para resolver cada caso del problema. Usar la misma semilla en los algoritmos para resolver cada caso del problema.

Cuanto menor es el valor de *Desv* para un algoritmo, mejor calidad tiene dicho algoritmo. Cuanto menor es el valor de *Tiempo*, mejor calidad tiene dicho algoritmo. Pero los métodos obtienen soluciones de la misma calidad (tienen valores de *Desv* similares) pero uno será mejor que el otro si emplea menos tiempo en media. En la web de la asignatura hay un ejemplo de cómo interpretar los resultados. Nuestros resultados deben estar un poco por debajo del óptimo.

La hoja Excel *Tablas_MDP_2020-21.xls* ya mencionada permite generar de forma automática las tablas de resultados para las tablas de resultados de la práctica.

Al profesor le sale 9 y pico de de

Debe tardar unos 5 minutos más o menos to

la función de fitness se va

5. Tablas de Resultados a Obtener

Se diseñará una tabla para cada algoritmo (*Greedy*, BL) donde se recojan los resultados de la ejecución de dicho algoritmo al conjunto de casos del problema. Tendrá la misma estructura que la Tabla 5.1.

Tabla 5.1: Resultados obtenidos por el algoritmo X en el MDP

Algoritmo X		
Caso	Desv	Tiempo
MDG-a_1_n500_m50	x	x
MDG-a_2_n500_m50	x	x
MDG-a_3_n500_m50	x	x
MDG-a_4_n500_m50	x	x
MDG-a_5_n500_m50	x	x
MDG-a_6_n500_m50	x	x
MDG-a_7_n500_m50	x	x
MDG-a_8_n500_m50	x	x
MDG-a_9_n500_m50	x	x
MDG-a_10_n500_m50	x	x
MDG-b_21_n2000_m200	x	x
MDG-b_22_n2000_m200	x	x
MDG-b_23_n2000_m200	x	x
MDG-b_24_n2000_m200	x	x
MDG-b_25_n2000_m200	x	x
MDG-b_26_n2000_m200	x	x
MDG-b_27_n2000_m200	x	x
MDG-b_28_n2000_m200	x	x
MDG-b_29_n2000_m200	x	x
MDG-b_30_n2000_m200	x	x
MDG-c_1_n3000_m300	x	x
MDG-c_2_n3000_m300	x	x
MDG-c_8_n3000_m400	x	x
MDG-c_9_n3000_m400	x	x
MDG-c_10_n3000_m400	x	x
MDG-c_13_n3000_m500	x	x
MDG-c_14_n3000_m500	x	x
MDG-c_15_n3000_m500	x	x
MDG-c_19_n3000_m600	x	x
MDG-c_20_n3000_m600	x	x

Finalmente, se construirá una tabla de resultados global que recoja los resultados medios de calidad y tiempo para todos los algoritmos considerados, tal como se muestra en la tabla 5.2.

Tabla 5.2: Resultados globales en el MDP

Algoritmo	Desv	Tiempo
<i>Greedy</i>	x	x
BL	x	x

A partir de los datos mostrados en estas tablas, el estudiante realizará un análisis de los resultados obtenidos, **que influirá significativamente en la calificación de la práctica**. En dicho análisis se deben comparar los distintos algoritmos en términos de calidad de las soluciones y tiempo requerido para producirlas. Por otro lado, se puede analizar también el comportamiento de los algoritmos en algunos de los casos individuales que presenten un comportamiento más destacado.

6. Documentación y Ficheros a Entregar

En general, la **documentación** de ésta y de cualquier otra práctica será un fichero pdf que deberá incluir, al menos, el siguiente contenido:

- Portada con el número y título de la práctica, el curso académico, el nombre del problema escogido, los algoritmos considerados; el nombre, DNI y dirección e-mail del estudiante, y su grupo y horario de prácticas.
- Índice del contenido de la documentación con la numeración de las páginas.
- Breve** descripción/formulación del problema (**máximo 1 página**). Podrá incluirse el mismo contenido repetido en todas las prácticas presentadas por el estudiante.
- Breve descripción de la aplicación de los algoritmos empleados al problema (**máximo 4 páginas**): Todas las consideraciones comunes a los distintos algoritmos se describirán en este apartado, que será previo a la descripción de los algoritmos específicos. Incluirá por ejemplo la descripción del esquema de representación de soluciones y la **descripción en pseudocódigo (no código)** de la función objetivo y los operadores comunes.
- Descripción en **pseudocódigo** de la **estructura del método de búsqueda** y de todas aquellas **operaciones relevantes** de cada algoritmo. Este contenido, específico a cada algoritmo se detallará en los correspondientes guiones de prácticas. El pseudocódigo **deberá forzosamente reflejar la implementación/ el desarrollo realizados** y no ser una descripción genérica extraída de las transparencias de clase o de cualquier otra fuente. La descripción de cada algoritmo no deberá ocupar más de **2 páginas**.

Para esta primera práctica, se incluirán al menos las descripciones en pseudocódigo de:

- Para el algoritmo BL, el métodos de exploración del entorno, el operador de generación de vecino, la factorización de la BL y la generación de soluciones aleatorias.
- Descripción en **pseudocódigo** del algoritmo de comparación, el *Greedy*.
 - Breve explicación del **procedimiento considerado para desarrollar la práctica**: implementación a partir del código proporcionado en prácticas o a

partir de cualquier otro, o uso de un framework de metaheurísticas concreto. Inclusión de un pequeño **manual de usuario describiendo el proceso para que el profesor de prácticas pueda replicarlo**.

- h) Experimentos y análisis de resultados:
 - Descripción de los casos del problema empleados y de los valores de los parámetros considerados en las ejecuciones de cada algoritmo (**incluyendo las semillas utilizadas**).
 - Resultados obtenidos según el formato especificado.
 - Análisis de resultados. El análisis deberá estar orientado a **justificar** (según el comportamiento de cada algoritmo) **los resultados** obtenidos en lugar de realizar una mera “lectura” de las tablas. Se valorará la inclusión de otros elementos de comparación tales como gráficas de convergencia, boxplots, análisis comparativo de las soluciones obtenidas, representación gráfica de las soluciones, etc.
- i) Referencias bibliográficas u otro tipo de material distinto del proporcionado en la asignatura que se haya consultado para realizar la práctica (en caso de haberlo hecho).

Aunque lo esencial es el contenido, también debe cuidarse la presentación y la redacción. **La documentación nunca deberá incluir listado total o parcial del código fuente.**

En lo referente al **desarrollo de la práctica**, se entregará una carpeta llamada **software** que contenga una versión ejecutable de los programas desarrollados, así como el código fuente implementado o los ficheros de configuración del framework empleado. El código fuente o los ficheros de configuración se organizarán en la estructura de directorios que sea necesaria y deberán colgar del directorio FUENTES en el raíz. Junto con el código fuente, hay que incluir los ficheros necesarios para construir los ejecutables según el entorno de desarrollo empleado (tales como *.prj, makefile, *.ide, etc.). La versión ejecutable de los programas se incluirá en un subdirectorio del raíz de nombre BIN. En este mismo directorio se adjuntará un pequeño fichero de texto de nombre LEEME que contendrá breves reseñas sobre cada fichero incluido en el directorio. Es importante que los programas realizados puedan leer los valores de los parámetros de los algoritmos desde fichero, es decir, que no tengan que ser recompilados para cambiar éstos ante una nueva ejecución. Por ejemplo, la semilla que inicializa la secuencia pseudoaleatoria debería poder especificarse como un parámetro más.

El fichero pdf de la documentación y la carpeta software serán comprimidos en un fichero .zip etiquetado con los apellidos y nombre del estudiante (Ej. Pérez Pérez Manuel.zip). Este fichero será entregado por internet a través del espacio de la asignatura en PRADO.

7. Método de Evaluación

Tanto en esta práctica como en las siguientes, se indicará la puntuación máxima que se puede obtener por cada algoritmo y su análisis. **La inclusión de trabajo voluntario** (desarrollo de variantes adicionales, experimentación con diferentes parámetros, prueba con otros operadores o versiones adicionales del algoritmo, análisis extendido, etc.) **podrá incrementar la nota final** por encima de la puntuación máxima definida inicialmente.

En caso de que el comportamiento del algoritmo en la versión implementada/ desarrollada no coincida con la descripción en pseudocódigo o no incorpore las componentes requeridas, se podría reducir hasta en un 50% la calificación del algoritmo correspondiente.