

Inteligencia Artificial Distribuida
Proyecto Compra Genética
Prof. Carlos B. Ogando M.

PROYECTO COMPRA GENÉTICA

HONESTIDAD ACADÉMICA

Como de costumbre, se aplica el código de honor estándar y la política de probidad académica. Las presentaciones isomórficas a (1) las que existen en cualquier lugar en línea, (2) las enviadas por sus compañeros de clase, o (3) las enviadas por los estudiantes en semestres anteriores serán consideradas plagio.

INSTRUCCIONES

En esta tarea, crearán un modelo de IA empleando algoritmos genéticos que permita optimizar las compras en supermercados (y otros tipos de tienda).

Lea atentamente todas las secciones de las instrucciones.

I. Introducción

II. Algoritmos genéticos

III. Planteamiento del problema

IV. Qué necesita enviar

I. Introducción

Los supermercados son catedrales del consumismo. Son entornos de marketing perfectamente perfeccionados, que se benefician de millones de libras de investigación sobre cómo animarnos y seducirnos para que compremos y gastemos más de lo que deberíamos. Esto significa que, como consumidores, debemos aprender a contraatacar.

Si quieres enseñarle a un niño de ocho años sobre el dinero, el mejor lugar para empezar es un supermercado. Pregúnteles qué pueden oler. Generalmente será pan o una panadería, ya que el olor nos da hambre y es probable que compremos más comida, por lo que el supermercado se beneficia.

¿Qué tal si hacemos un modelo de IA que permita optimizar estas costosas compras? Podemos crear un sistema que empleando algoritmos genéticos encuentra la mejor ruta para comprar productos de una lista que tengamos en función de los catálogos de las distintas sucursales con sus precios y otras variables como los costos de transporte.

II. Algoritmos genéticos

Los algoritmos genéticos (AG) funcionan entre el conjunto de soluciones de un problema llamado fenotipo, y el conjunto de individuos de una población natural, codificando la información de cada solución en una cadena, generalmente binaria, llamada cromosoma. Los símbolos que forman la cadena son llamados genes. Cuando la representación de los cromosomas se hace con cadenas de dígitos binarios se le conoce como genotipo. Los cromosomas evolucionan a través de iteraciones, llamadas generaciones. En cada generación, los cromosomas son evaluados usando alguna medida de aptitud. Las siguientes generaciones (nuevos cromosomas), son generadas aplicando los operadores genéticos repetidamente, siendo estos los operadores Los algoritmos genéticos son de probada eficacia en caso de querer calcular funciones no derivables (o de derivación muy compleja) aunque su uso es posible con cualquier función.

Deben tenerse en cuenta también las siguientes consideraciones:

Si la función a optimizar tiene muchos máximos/mínimos locales se requerirán más iteraciones del algoritmo para "asegurar" el máximo/mínimo global.

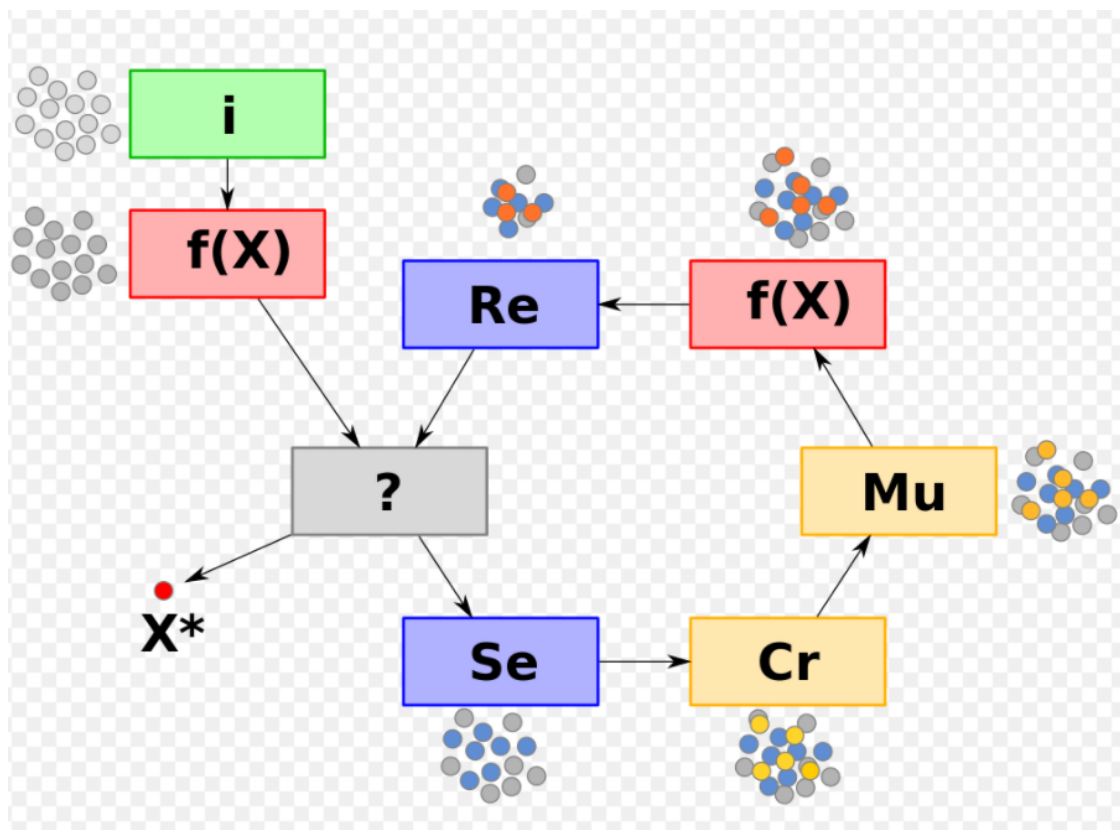
Si la función a optimizar contiene varios puntos muy cercanos en valor al óptimo, solamente podemos "asegurar" que encontraremos uno de ellos (no necesariamente el óptimo). de selección, cruzamiento, mutación y reemplazo.

Un algoritmo genético puede presentar diversas variaciones, dependiendo de cómo se decide el reemplazo de los individuos para formar la nueva población. En general, el pseudocódigo consiste de los siguientes pasos:

- **Inicialización:** Se genera aleatoriamente la población inicial, que está constituida por un conjunto de cromosomas los cuales representan las posibles soluciones del problema. En caso de no hacerlo aleatoriamente, es importante garantizar que dentro de la población inicial, se tenga la diversidad estructural de estas soluciones para tener una representación de la mayor parte de la población posible o al menos evitar la convergencia prematura.
- **Evaluación:** A cada uno de los cromosomas de esta población se aplicará la función de aptitud para saber cómo de "buena" es la solución que se está codificando.
- **Condición de término:** El AG se deberá detener cuando se alcance la solución óptima, pero esta generalmente se desconoce, por lo que se deben utilizar otros criterios de detención. Normalmente se usan dos criterios: correr el AG un número máximo de iteraciones (generaciones) o detenerlo cuando no haya cambios en la población. Mientras no se cumpla la condición de término se hace lo siguiente:
 - **Selección:** Después de saber la aptitud de cada cromosoma se procede a elegir los cromosomas que serán cruzados en la siguiente

generación. Los cromosomas con mejor aptitud tienen mayor probabilidad de ser seleccionados.

- **Recombinación o cruzamiento:** La recombinación es el principal operador genético, representa la reproducción sexual, opera sobre dos cromosomas a la vez para generar dos descendientes donde se combinan las características de ambos cromosomas padres.
- **Mutación:** Modifica al azar parte del cromosoma de los individuos, y permite alcanzar zonas del espacio de búsqueda que no estaban cubiertas por los individuos de la población actual.
- **Reemplazo:** Una vez aplicados los operadores genéticos, se seleccionan los mejores individuos para conformar la población de la generación siguiente.



(!) LEER ARTÍCULO DE WIKIPEDIA EN LA SECCIÓN “**INTRODUCCIÓN**” PARA CONOCER LOS PASOS PARA APLICAR EL ALGORITMO, FÓRMULAS Y CÓMO APLICARLO.

https://es.wikipedia.org/wiki/Algoritmo_gen%C3%A9tico

(!) VER EL VIDEO SIGTE. PARA UNA EXPLICACIÓN DETALLADA DEL ALGORITMO.

<https://www.youtube.com/watch?v=HeG2cUp0TVY>

Obtenido gracias a Wikipedia

III. Planteamiento del problema

La optimización de la compra en supermercados resulta ser un problema bastante interesante de optimización.

3.1. Objetivo

El objetivo de este proyecto es diseñar un sistema de optimización de compras en supermercados que minimice el **costo total de la compra**, incluyendo tanto el **precio de los productos como el gasto en combustible**. Utilizando algoritmos genéticos, los estudiantes desarrollarán una solución que determine la ruta más eficiente para adquirir **todos los productos necesarios**, considerando múltiples sucursales de supermercados, cada una con su propio catálogo de productos y distancias específicas **desde la casa del comprador**.

3.2. Descripción del Problema

Sucursales y catálogo de productos:

- Se dispone de un número determinado de sucursales de supermercados.
- Cada sucursal tiene un **catálogo de productos**, donde se especifican los productos disponibles y sus respectivos precios.

Distancias:

- Se proporciona la distancia entre cada sucursal y la casa del comprador.
- La distancia entre **las diferentes sucursales también es conocida**.

Otros parámetros:

- Se cuenta con una lista de productos que el comprador necesita adquirir.
- **Se cuenta con el precio de combustible por distancia.**

III. Qué necesita enviar

Su trabajo está en desarrollar un sistema con un modelo de IA que permita hallar la ruta de compra óptima en supermercados dada una lista de compra empleando **algoritmos genéticos**.

Para completar satisfactoriamente este proyecto es necesario cumplir los siguientes requerimientos:

PARTE 1: Diseño de la solución

Entregar y presentar un documento que explique cómo paso a paso empleando algoritmos genéticos se puede resolver este problema, con una demostración matemática paso a paso.

PARTE 2: Implementación

- **Objetivo:** Minimizar el costo de productos y transporte **que asumirá una persona para comprar X productos**.

- **Optimización de costo de productos.**
 - Se definirá una lista de sucursales donde venden los productos.
 - Cada sucursal tiene una lista específica de productos que vende cada uno con su precio.
 - Cada sucursal está ubicada en un grafo (que se debe convertir a totalmente conectado). El grafo después de convertirse en totalmente conectado tendrá su distancia mínima a todas las demás sucursales y al punto de partido de la persona target.
- **Optimización de transporte.**
 - Las variables para la optimización de transporte son el precio del combustible por distancia y la distancia de los nodos del grafo.
- **Entradas y salidas:**

INPUT: Recibe un Excel con el catálogo de sucursales y lista de aristas del grafo. Recibe una lista de compras y un costo por combustible.

OUTPUT: Retorna la ruta óptima de compra indicando en orden de recorrido cuál sucursal visitar y qué comprar. Ejemplo:

RESPUESTA> (Bravo, [Arroz, Aceite]) -> (La Sirena, Cerveza)

Los siguientes componentes específicamente deben estar:

- Clase que empotra las funciones del algoritmo genético implementado (una función por cada paso).
- Estos algoritmos deben estar adaptados al problema de optimización seleccionado en términos de funciones de fitness.
- Estas funciones de las clases deben estar parametrizadas en base a los hiperparámetros de cada algoritmo. Si usa heurísticas debe crear una función en el código también (si la heurística lo amerita).
- Su programa debe recibir un archivo Excel con dos hojas:
 - La primera hoja debe tener el catálogo de productos de cada sucursal con su precio. (para construir la matriz de precio)
 - La segunda debe tener las aristas y sus tamaños (para construir el grafo [matriz de distancia])
- Su programa debe recibir como parámetro:
 - Lista de productos que se quiere procesar.
 - Precio de combustible
- Definir una sintaxis coherente para los genes de los pobladores.
- Seleccionar funciones de fitness, crossover, mutación y selección coherente para la sintaxis de genes definidos.
- Debe probar al menos con:
 - 2 técnicas de crossover.
 - 2 técnicas de mutación.
 - 2 técnicas de selección.


PARTE 3: Validación

Para validar se le proporcionará un dataset con diferentes problemas de optimización de compra y su solución. De esta forma podrá saber qué tan bien funciona su algoritmo.

En la carpeta con la documentación del proyecto encontrarán también el problemset de validación y los input de cada problema.

La carpeta aparte del PDF del proyecto debe incluir:


 in


 problemset


El dataset con los problemas llamado “problemset.xlsx” contiene varias columnas:


Nombre de la columna	Descripción
problem_cluster	Tipo de problema
catalog	Ruta del archivo que contiene el catálogo y el grafo
fuel_price	Precio del combustible por unidad de distancia
query	Lista de compras del usuario
qbf best solution	Mejor solución por quasi-fuerza bruta
qbf best value	Mejor costo por quasi-fuerza bruta
ga best solution	Mejor solución por algoritmos genéticos
ga best value	Mejor costo por algoritmos genéticos
gc best solution	Mejor solución por greedy-más barato
gc best value	Mejor costo por greedy-más barato
gn best solution	Mejor solución por greedy- más cercano
gn best value	Mejor costo por greedy- más cercano


Para cada problema se encontrarán los inputs del mismo en la carpeta “in”


 big-100s-300p-f0_10-q1_20-random_catalog


 medium-20s-50p-f0_10-q1_20-random_catalog


 medium-30s-80p-f0_10-q1_20-random_catalog


 small-3s-4p-f1_10-q1_4-random_catalog


 small-5s-10p-f0_10-q1_10-jumbocheap_catalog

 small-5s-10p-f0_10-q1_10-parque_catalog

 small-5s-10p-f0_10-q1_10-punico_catalog

 small-5s-10p-f0_10-q1_10-random_catalog

 small-5s-10p-f0_10-q1_10-same_dist_catalog

 small-5s-10p-f0_10-q1_10-same_price_catalog

Esta carpeta contiene los excels con el catálogo de productos de las tiendas en la primera hoja “catalogo” y el grafo en la segunda hoja “grafo”.

store	product	price
La Sirena	Salsa ranchera	100.99
Bravo	Toallas sanitarias	100.99
La Sirena	Whisky	100.99
Bravo	Antihistamínicos	100.99
Super Pola	Toallas sanitarias	100.99
Super Pola	Antihistamínicos	100.99
Supermercado Nacional	Whisky	100.99
Supermercado Nacional	Pulseras	100.99
Bravo	Aceitunas verdes	100.99
Super Pola	Suavizante de telas	100.99
Supermercado Nacional	Antihistamínicos	100.99
Jumbo	Aceitunas verdes	100.99
Jumbo	Toallas sanitarias	100.99
Jumbo	Salsa ranchera	100.99
Supermercado Nacional	Consolas de videojuegos	100.99
Bravo	Reproductores de DVD	100.99
Bravo	Pulseras	100.99
Bravo	Whisky	100.99

< > catalogo grafo +

edge 1	edge 2	distance
Casa	Super Pola	1
Jumbo	Super Pola	15
Casa	Casa	19
Bravo	La Sirena	6
Jumbo	Supermercado Nacional	13
La Sirena	Jumbo	15
Bravo	Super Pola	4
Super Pola	Super Pola	14
Super Pola	Supermercado Nacional	17
Supermercado Nacional	Casa	8
Supermercado Nacional	Jumbo	18
La Sirena	Super Pola	12
Supermercado Nacional	Bravo	10
Super Pola	La Sirena	13
Super Pola	Jumbo	4
La Sirena	Casa	5
Jumbo	Casa	5
Bravo	Casa	0

< > catalogo grafo +

La casa del usuario tendrá el nombre “Casa” en todos los archivos.

PARTE 4: Documentación

- Documentación de las fuentes del proyecto con descripción de los métodos y clases desarrollados, parámetros de entrada, salida, hiperparámetros, etc.
 - Describir sintaxis de genes.
 - Describir todas las funciones genéticas (fitness, selección, crossover, mutación...)
 - Definir riesgos y limitaciones del proyecto.
 - Oportunidades de mejoras.
 - Rendimiento de los resultados y tiempos de ejecución.
 - Escalabilidad.
- Documento con el análisis comparativo (benchmark) de las estadísticas en términos de tiempo, memoria, número de iteraciones, probabilidad de mutación, tasa de selección, técnicas usadas, entre otros parámetros usados en el proyecto para la optimización usando su algoritmo y los de referencia como benchmark.
- Realizar benchmark con la data de validación usando las siguientes métricas y documentar hallazgos y conclusiones.
 - Error (E)
 - Error absoluto (AE)
 - Error relativo (RE)
 - Error absoluto promedio (MAE)
 - Error cuadrado promedio (MSE)
 - Entre otros que sean de interés (pueden ser de regresión).
- Diagrama de flujo de datos del proyecto.
- Manual de usuario del sistema.

NOTA:

Este proyecto **debe ser explicado en clase** en la fecha indicada por el maestro, **de lo contrario se perderán todos los puntos** del mismo.