**PROBLEMA DE RUTEO DE VEHÍCULOS DE DOS ESCALONES CON VENTANAS DE TIEMPO**

El 2E-CVRP es presentado de manera formal por González en el 2007, este problema puede verse como una extensión del CVRP o como un caso específico del 2E-LRP en el que la ubicación de las instalaciones es conocida de antemano, por lo tanto la localización de las instalaciones (depósito central, centros de distribución y ubicación de los clientes) forman parte de los parámetros del problema. Autores como Chaug-Ing Hsu, Sheng-Feng Hung, Hui-Chieh Li, en el 2007 ampliaron el problema al presentar un modelo de enrutamiento de vehículos con ventanas de tiempo para la entrega de alimentos perecederos (SVRPTW), dicha restricción surgió debido al deterioro de los alimentos con una vida útil corta, tales como la leche, frutas y verduras, deterioro que ocasiona una pérdida en los ingresos de los minoristas, la cual debería ser transferida a los operadores de los centros de distribución como penalización por retrasos en la entrega.

El 2E-CVRPTW consta de dos niveles o escalones, y por lo tanto de dos tipos de rutas, es preciso resaltar que el modelo matemático se enfoca en solucionar el ruteo para los dos escalones de forma integral, a la vez que se considera minimizar el costo del problema, el cual está asociado a la distancia de los recorridos de las rutas en ambos niveles y el costo por la utilización de los vehículos de cada nivel; se considera como ruta de primer nivel a una que conecta un depósito central con uno o más centros de distribución conocidos como satélites y como ruta de segundo nivel a una que conecta a un satélite con uno o más clientes los cuales requieren una demanda de productos. Cada cliente debe ser servido únicamente por un vehículo del segundo nivel, las rutas de segundo nivel debe comenzar y terminar en el mismo satélite, mientras que para las de primer nivel no se requiere una asignación, debido a que estas parten del depósito central. La demanda de una instalación satelital es la suma de la demanda de los clientes que se le asignan, por lo tanto, cualquier cambio en la asignación del cliente afecta el enrutamiento del primer nivel. Los vehículos de primer nivel pueden transportar la carga de uno o más clientes así como servir más de un satélite en la misma ruta, cuando la demanda de un satélite supera la capacidad de los vehículos de primer nivel estos pueden ser visitados por más de un vehículo. Usualmente las restricciones de capacidad son consideradas tanto para los satélites como para los vehículos. Se considera una flota de vehículos con capacidad homogénea para cada uno de los niveles, siendo la flota del primer nivel de mayor capacidad con respecto a la del segundo. Cada cliente impone una ventana de tiempo, la cual se debe cumplir para que se pueda realizar la entrega, lo que determina unas ventanas de tiempo duras.

**DESCRIPCION DETALLADA DEL 2E-CVRPTW**

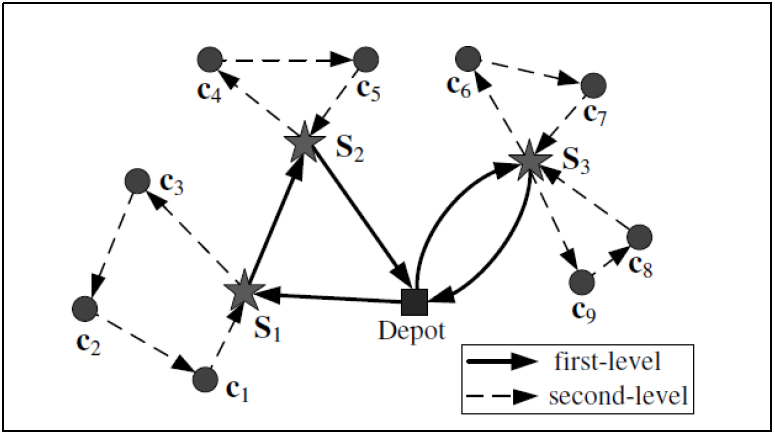
De acuerdo con las especificaciones planteadas por Zeng, el 2E-VRP se puede definir en un grafo ponderado, completo y no dirigido G= (V, A) con un conjunto de nodos V y un conjunto de arcos A. El conjunto V está compuesto por tres tipos de nodos: un depósito central , un subconjunto , con satélites y un subconjunto de clientes.

Respecto al conjunto de arcos , compuesto por subconjuntos de arcos y que representan una ruta que conecta los nodos para el primer y segundo nivel respectivamente, al cual se le asocia un costo por su recorrido ; en el que estos satisfacen la desigualdad triangular en el primer y segundo nivel.

A cada cliente está asociado a una demanda , la cual no puede ser dividida entre los vehículos del segundo nivel, la demanda no puede ser entregada directamente desde el depósito central a los clientes, esta debe ser consolidada en alguno de los satélites para su posterior envío al cliente. Sin embargo, las entregas a los satélites en el primer nivel si pueden ser divididas.

Cada ruta del segundo nivel debe comenzar y terminar en el mismo satélite. Los vehículos tienen una capacidad limitada que debe ser respetada, esta capacidad es la misma para todos los vehículos pertenecientes al mismo nivel, pero puede diferir para cada nivel. Las capacidades en los vehículos para el primer y segundo nivel están denotadas por y respectivamente. Adicional a esto, cada satélite tiene una capacitad limitada la cual restringe la cantidad que llega a este desde las rutas de primer nivel. La demanda de cada satélite es la demanda total de los clientes asignados a él, por lo que cada satélite debe recibir suficiente carga del depósito para satisfacer a los clientes del segundo nivel.

No se introduce ninguna limitación adicional en el tamaño de la ruta, ni en la longitud, ni en el número de clientes visitados. Sin embargo, no se permite la activación del servicio si un vehículo llega antes o después de los límites inferiores y superiores de cada cliente, implementando ventanas de tiempo duras. El objetivo del 2E-CVRPTW es minimizar los costos de recorrido sumado con los costos por la utilización de los vehículos.



*Figura 1: Ilustración del 2E-VRP*

**FORMULACIÓN MATEMATICA DEL 2E-CVRPTW**

Acorde a la descripción previa del 2E-CVRPTW se introduce el modelo matemático. Las anotaciones utilizadas en este documento se adopta principalmente de Niño Sáenz, A. (2017).

**Grupos**

Deposito central.

Conjunto de satélite.

Conjunto de clientes.

: Conjunto de vehículos de primer nivel .

: Conjunto de vehículos de segundo nivel .

Conjunto de arcos.

Subconjunto de arcos atravesados por un vehículo de primer nivel

Subconjunto de arcos atravesados por un vehículo de segundo nivel

: Conjunto de vértices.

**Parámetros**

: Costo por recorrido para el primer nivel .

: Costo por minuto del vehículo para el segundo nivel

: Costo por utilizar vehículos .

: Costo por utilizar vehículos .

Número de clientes.

Número de satélites.

: Capacidad de un vehículo de primer nivel.

: Capacidad de un vehículo de segundo nivel.

: Demanda del cliente .

: Capacidad del satélite

: Límite inferior ventana de tiempo cliente

: Límite inferior ventana de tiempo cliente

Tiempo de visita en el nodo .

: Hora de inicio de los satélites.

: Hora de cierre de los satélites.

Distancia de los arcos (i, j) .

Distancia de los arcos (i, j) .

Tiempo de viaje para los arcos (i, j) .

Constante muy grande.

**Variables**

: Variable binaria que es igual a 1 si el arco (i, j) es recorrido por el vehículo está en la solución e igual a cero en el caso contrario.

: Variable binaria que toma el valor de 1 si el arco (i, j) esrecorrido por el vehículo e igual a cero en el caso contrario.

Variable binaria que es igual a 1 si el vehículo h es utilizado e igual a cero en caso contrario.

Variable binaria que es igual a 1 si el vehículo p es utilizado e igual a cero en caso contrario.

Inicio de visita del vehículo en el cliente .

: Variable entera no negativa que representa la cantidad entregada por el vehículo h al satélite .

: Variable entera no negativa que representa la cantidad entregada por el vehículo p al cliente

(1)

Sujeto a

(2)

(3)

(4)

(5)

(6)

(7)

(8)

(9)

(10)

(11)

(12)

(13)

(14)

(15)

(16)

(17)

(18)

(19)

(20)

La función objetivo (1) minimiza el costo total de los dos escalones teniendo en cuenta la utilización de vehículos y el costo por recorrido. Las restricciones (2) y (3) aseguran que cada cliente debe ser visitado por exactamente un vehículo de segundo nivel y que se parte de este hacía otro nodo. Restricciones (4) y (5) limitan el número de rutas por vehículo a una, y caracterizan el flujo que deben seguir la flota de segundo nivel; la primera hace referencia al número de veces que un vehículo sale del satélite y la segunda, al número de veces que llega. (6) asegura que a cada cliente solo llegue un vehículo y el mismo salga de él; lo que permite que en el modelo no se formen ciclos. La restricción (7) asegura que el vehículo en el segundo nivel no pueda comenzar el servicio antes del inicio de la ventana de tiempo del cliente. (8) y (9) garantizan que el inicio del servicio en el cliente se lleve a cabo dentro de la ventana de tiempo correspondiente.

La restricción (10) permite que cada cliente sea servido dentro del intervalo asociado a los nodos de llegada y salida del satélite. La suma de las demandas de los clientes de una ruta no puede exceder la capacidad de los vehículos de segundo nivel y está dada por la restricción (11). Las restricciones (12) y (16) controlan el costo fijo de los vehículos de segundo y primer nivel. La restricción (13) establece que la capacidad de cada satélite debe ser mayor o igual a las demandas de los clientes asignados a este. La cantidad de demanda que llega desde las rutas de primer nivel al satélite debe ser igual a la suma de la demanda de los clientes asignados a él (14). La suma de las demandas de los satélites de una ruta no puede exceder la capacidad de los vehículos de primer nivel y está dada por la restricción (15). Las variables del modelo son de tipo binario y entero (17), (18), (19) y (20).

La función objetivo del problema se expresa como:

En la que se busca minimizar los costos de recorrido asociados a las rutas de primer nivel, segundo nivel, sumado al costo de utilización de los vehículos de cada uno de los niveles.

Respecto a las restricciones se tiene que:

* Cada cliente debe ser visitado por un solo vehículo el cual efectúa una única ruta de segundo nivel.
* Asegura que desde cada cliente se parta hacia otro nodo, una única vez por solo un vehículo.
* Limita el número de rutas por vehículo a una y caracteriza el flujo que debe seguir, haciendo referencia al número de veces que un vehículo sale del satélite.
* Limita el número de rutas por vehículo a una y caracteriza el flujo que debe seguir, haciendo referencia al número de veces que un vehículo llega al satélite.
* Asegura que no se formen ciclos en el modelo, de manera que solo llegue un vehículo al cliente y el mismo salga de él.
* Asegura que el vehículo de segundo nivel p no pueda comenzar el servicio, si la suma del tiempo de viaje de i a j, la duración del servicio en i y el tiempo total acumulado al inicio del servicio en i es mayor que la ventana de tiempo de j o del cliente siguiente.
* Garantiza que el inicio de servicio en el cliente i se lleve a cabo después del límite inferior de la ventana de tiempo correspondiente.
* Garantiza que el inicio de servicio en el cliente i se lleve a cabo después del límite superior de la ventana de tiempo correspondiente.
* Permite que cada cliente sea servido dentro del intervalo asociado a los nodos de llegada y salida del satélite, además condiciona el tiempo total de viaje de cada vehículo.
* La suma de las demandas de los clientes de una ruta no puede exceder la capacidad de los vehículos de segundo nivel.
* Los costos fijos de los vehículos de segundo nivel se controlan con la utilización de dichos vehículos y el recorrido de estos.
* Establece que la capacidad de cada satélite debe ser mayor o igual a las demandas de los clientes asignados a este.
* Las unidades que llegan desde las rutas de primer nivel al satélite debe ser igual a la suma de la demanda de los clientes asignados a él.
* La suma de las demandas de los satélites de una ruta no puede exceder la capacidad de los vehículos de primer nivel.
* Los costos fijos de los vehículos de primer nivel se controlan con la utilización de dichos vehículos y el recorrido de estos.
* Las variables del modelo son de tipo binario y entero.

**METAHEURÍSTICA APLICADA AL 2E-CVRPTW**

**PROCEDIMIENTO DE ALGORITMO GENETICO**

Son métodos adaptativos que pueden usarse para resolver problemas de búsqueda y optimización. Están basados en el proceso genético de los organismos vivos, capaces de ir creando soluciones para problemas del mundo real. A lo largo de las generaciones, las poblaciones evolucionan en la naturaleza de acorde con los principios de la selección natural y la supervivencia de los más fuertes postulados por Darwin (1859).

En la naturaleza los individuos de una población compiten entre sí en la búsqueda de recursos tales como comida, agua y refugio. Aquellos individuos que tienen más éxito en sobrevivir y en atraer compañeros tienen mayor probabilidad de generar un gran número de descendientes; por el contrario, individuos poco dotados producirán un menor número de descendientes. Esto significa que los genes de los individuos mejor adaptados se propagarán en sucesivas generaciones, logrando una evolución de las especies con características cada vez mejor adaptadas al entorno en el que viven, la combinación de buenas características provenientes de diferentes ancestros, puede a veces producir descendientes “súper individuos”, cuya adaptación es mucho mayor que la de cualquiera de sus ancestros.

El poder de los algoritmos genéticos proviene del hecho de que se trata de una técnica robusta y pueden tratar con éxito una gran variedad de problemas provenientes de diferentes áreas, usando una analogía directa con el comportamiento natural. Trabajan con una población de individuos, cada uno de los cuales representa una solución factible a un problema dado, se le asigna a cada individuo un valor o puntuación relacionada con la bondad de dicha solución. En la naturaleza esto equivaldría al grado de efectividad de un organismo para competir por unos determinados recursos, cuanto mayor sea la adaptación de un individuo al problema, mayor será la probabilidad de que el mismo sea seleccionado para reproducirse, cruzando su material genético con otro individuo seleccionado de igual forma. Este cruce producirá nuevos individuos descendientes de los anteriores, los cuales comparten algunas de las características de sus padres; cuanto menor sea la adaptación de un individuo, menor será la probabilidad de que dicho individuo sea seleccionado para la reproducción, y por tanto de que su material genético se propague en sucesivas generaciones.

De esta manera se produce una nueva población de posibles soluciones, la cual reemplaza a la anterior y verifica la interesante propiedad de que contiene una mayor proporción de buenas características en comparación con la población anterior. Así a lo largo de las generaciones las buenas características se propagan a través de la población, favoreciendo el cruce de los individuos mejor adaptados, van siendo exploradas las áreas más prometedoras del espacio de búsqueda.

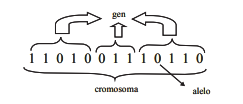
Durante la ejecución del algoritmo, los padres deben ser seleccionados para la reproducción, a continuación dichos padres seleccionados se cruzaran generando hijos, sobre cada uno de los cuales actuara un operador de mutación. El resultado de la combinación de las anteriores funciones será un conjunto de individuos (posibles soluciones al problema), los cuales en la evolución del algoritmo genético formaran parte de la solución.

**CODIFICACIÓN DE PROBLEMAS**

Cualquier solución potencial a un problema puede ser presentada dando valores a una serie de parámetros. El conjunto de todos los parámetros (genes en la terminología de algoritmo genético) se codifica en una cadena de calores denominados cromosomas.

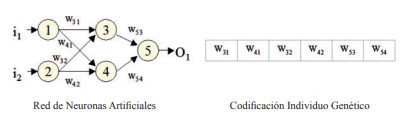
El conjunto de parámetros representado por un cromosoma particular recibe el nombre de genotipo, que contiene la información necesaria para la construcción del organismo, es decir, la solución real del problema, denominada fenotipo. Por ejemplo, en términos biológicos, la información genética contenida en el ADN de un individuo seria el genotipo, mientras que la expresión de ese ADN (el propio individuo) sería el fenotipo.

Desde los primeros estudios la codificación suele hacerse mediante valores binarios, se asigna un determinado número de bits a cada parámetro y se realiza una discretización de la variable representada por cada gen. El número de bits asignados dependerá del grado de ajuste que se desee alcanzar, evidentemente no todos los parámetros tienen por qué estar codificados con el mismo número de bits. Cada uno de los bits pertenecientes a un gen suele recibir el nombre de alelo.



*Figura 2: Individuo genético binario*

Sin embargo, también existen representaciones que codifican directamente cada parámetro con un valor entero, real o en punto flotante. Un ejemplo típico de la aplicación de los algoritmos genéticos es representar una red de neuronas artificiales para posteriormente realizar el proceso de optimización de los pesos sinápticos; codificar una red de neuronas artificiales en forma de cromosomas es tan sencillo como asignar un gen del cromosoma a cada uno de los pesos de la red. También se podrían añadir genes que indiquen el número de capas y el número de elementos de procesado en cada una.



*Figura 3: Codificación de una red de neuronas artificiales*

Una generación se obtiene por medio de los operadores de reproducción. Existen dos tipos:

* ***Cruce:*** Se trata de una reproducción de tipo sexual, se genera una descendencia a partir del mismo número de individuos (generalmente 2) de la generación anterior.
* ***Copia:*** Se trata de una reproducción de tipo asexual, un número de individuos determinado pasa sin sufrir ninguna variación directamente a la siguiente generación.

Si desea optarse por una estrategia elitista, los mejores individuos de cada generación se copian siempre en la población temporal, para evitar su perdida.

A continuación comienza a generarse la nueva población en base a la aplicación de los operadores genéticos de cruce y/o copia. Una vez generados los nuevos individuos se realiza la mutación con una probabilidad *Pm*. La probabilidad de mutación suele ser muy baja, por lo general entre 0.5% y el 2%.

Se termina este proceso cuando se alcanzan algunos criterios como:

* Los mejores individuos de la población representan soluciones suficientemente buenas para el problema que se desea resolver.
* La población ha convergido. Un gen ha convergido cuando el 95% de la población tiene el mismo valor para él, en el caso de trabajar con codificaciones binarias, o valores dentro de un rango especificado en el caso de trabajar con otro tipo de codificaciones. Una vez que todos los genes alcanzan la convergencia se dice que la población ha convergido. Cuando esto ocurre la media de bondad de la población se aproxima a la bondad del mejor individuo.
* Se ha alcanzado el número de generaciones máximo especificado.

Sobre este algoritmo se han definido numerosas variantes, quizá una de las más extendidas consiste en prescindir de la población temporal de manera que los operadores genéticos de cruce y mutación se aplican directamente sobre la población genética. Con esta variante el proceso de cruce varia ligeramente, no basta en el caso de que el cruce se produzca, con insertar directamente la descendencia en la población, puesto que el número de individuos de la población se le ha de hacer sitio. Es decir, para ubicar a los descendientes generados previamente se han de eliminar otros individuos de la población genética.

Otra variación común consiste en la modificación del esquema de selección de los individuos que serán mutados; así como la opción habitual de selección aleatoria del individuo a mutar entre todos los que forman parte de la población genética.

**Operadores genéticos**

Para el paso de una generación a la siguiente se aplican una serie de operadores genéticos, los más empleados son los de selección, cruce, copia y mutación. En el caso de no trabajar con una población intermedia temporal también cobran relevancia los algoritmos de reemplazo.

1. *Selección*

Los algoritmos de selección serán los encargados de escoger que individuos van a disponer de oportunidades de reproducirse y cuáles no, es decir, están relacionados con su valor de ajuste. No se debe, sin embargo, eliminar por completo las opciones de reproducción de los individuos menos aptos, pues en pocas generaciones la población se volvería homogénea.

Estos algoritmos de selección pueden ser divididos en dos grandes grupos: *probabilísticos y determinísticos*, ambos tipos de algoritmo basan su funcionamiento en el principio indicado anteriormente. Sin embargo, el primer tipo adjudica estas posibilidades con un importante componente basado en el azar, donde se encuentran los algoritmos de selección por ruleta o por torneo; el segundo tipo engloba una serie de algoritmos que, dado el ajuste conocido de cada individuo, permite asignar a cada uno el número de veces que será escogido para reproducirse. Esto puede evitar problemas de predominancia de ciertos individuos y cada uno de estos algoritmos presenta variaciones respecto al número de veces que se tomaran los mejores y peores y, de esta forma, se impondrá una presión en la búsqueda en el espacio de estados en la zona donde se encuentra el mejor individuo, o bien que se tienda a repartir la búsqueda por el espacio de estados, pero sin dejar de tender a buscar en la mejor zona.

* 1. *Selección por ruleta*

Posiblemente el método más utilizado desde los orígenes de los algoritmos genéticos (Blickle & Thiele, 1995). A cada uno de los individuos de la población se le asigna una parte proporcional a su ajuste de una ruleta, de tal forma que la suma de todos los porcentajes sea la unidad. Los mejores individuos recibirán una porción de la ruleta mayor que la recibida por los peores, generalmente la población esta ordenada en base al ajuste, por lo que las proporciones más grandes se encuentran al inicio de la ruleta. Para seleccionar un individuo basta con generar un numero aleatorio del intervalo [0…1] y devolver el individuo situado en esa posición de la ruleta, la cual, se suele obtener recorriendo los individuos de la población y acumulando sus proporciones de ruleta hasta que la suma exceda el valor obtenido. Presenta además el inconveniente de que el peor individuo puede ser seleccionado más de una vez.

* 1. *Selección por torneo*

Consiste en escoger a los individuos genéticos en base a comparaciones directas entre sus genotipos. Existen dos versiones de selección mediante torneo, el torneo determinístico que selecciona al azar un numero p de individuos (generalmente se escoge p=2); de entre los individuos seleccionados se selecciona el más apto para pasarlo a la siguiente generación. El torneo probabilístico únicamente se diferencia en el paso de la selección del ganador del torneo. En vez de escoger siempre el mejor se genera un numero aleatorio del intervalo [0…1], si es mayor que un parámetro p (fijado para todo el proceso evolutivo) se escoge el individuo más apto y en caso contrario el menos apto. Generalmente p toma valores en el rango 0.5<p≤1.

Variando el número de individuos que participan en cada torneo se puede modificar la presión de selección, cuando participan muchos individuos en cada torneo, la presión de selección es elevada y los peores individuos apenas tienen oportunidades de reproducción, un caso particular es el elitismo global. Se trata de un torneo en el que participan todos los individuos de la población, con la cual la selección se vuelve totalmente determinística. Cuando el tamaño del torneo es reducido, la presión de selección disminuye y los peores individuos tienen más oportunidades de ser seleccionados.

Elegir uno u otro método se selección determinara la estrategia de búsqueda del algoritmo genético, si se opta por un método con una alta presión de selección se centra la búsqueda de las soluciones en un entorno próximo a las mejores soluciones actuales. Por el contrario, optando por una presión de selección menor se deja el camino abierto para la exploración de nuevas regiones del espacio de búsqueda.

1. *Cruce*

Una vez seleccionados los individuos, estos son recombinados para producir la descendencia que se insertara en la siguiente generación, su importancia para la transición entre generaciones es elevada puesto que las tasas de cruce con las que se suele trabajar rondan el 90%.

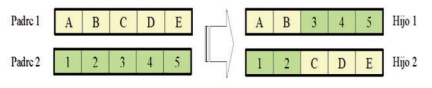
Los diferentes métodos de cruce podrán operar de dos formas diferentes: si se opta por una estrategia destructiva los descendientes se insertaran en la población temporal aunque sus padres tengan mejor ajuste, por el contrario, utilizando una estrategia no destructiva la descendencia pasara a la siguiente generación únicamente si supera la bondad del ajusta de los padres.

La idea principal del cruce se basa en que, si se toman dos individuos correctamente adaptados al medio y se obtiene una descendencia que comparta genes de ambos, existe la posibilidad de que los genes heredados sean precisamente los causantes de la bondad de los padres. Al compartir las características buenas de dos individuos, la descendencia, o al menos parte de ella, debería tener una bondad mayor que cada uno de los padres por separado. Si el cruce no agrupa las mejores características en uno de los hijos y la descendencia tiene un peor ajuste que los padres no significan que se esté dando un paso atrás. Optando por una estrategia de cruce no destructiva garantizamos que pasen a la siguiente generación los mejores individuos, si aun con un ajuste peor, se opta por insertar a la descendencia, y puesto que los genes de los padres continuaran en la población en posteriores cruces se podrán volver a obtener estos padres, recuperando así la bondad previamente perdida.

Existen multitud de algoritmos de cruce, los más empleados son:

* Cruce de 1 punto
* Cruce de 2 puntos
* Cruce uniforme
  1. *Cruce de 1 punto*

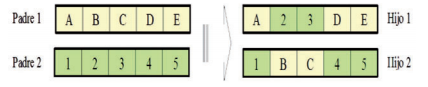
Una vez seleccionados dos individuos se cortan sus cromosomas por un punto seleccionado aleatoriamente para generar dos segmentos diferenciados en cada uno de ellos: la cabeza y la cola. Se intercambian las colas entre los dos individuos para generar los nuevos descendientes, de esta manera ambos descendientes heredan información genética de los padres.



*Figura 4: Cruce de 1 punto*

* 1. *cruce de 2 puntos*

Se trata de una generalización del cruce de 1 punto. En vez de cortar un único punto los cromosomas de los padres, como en el caso anterior, se realizan dos cortes. Se debe tener en cuenta que ninguno de estos puntos de corte coincida con el extremo de los cromosomas para garantizar que se originen tres segmentos. Para generar la descendencia se escoge el segmento central de uno de los padres y los segmentos laterales del otro padre.



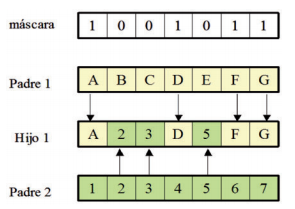
*Figura 5: Cruce de 2 puntos*

Generalizando, se pueden añadir más puntos de cruce dando lugar a algoritmos de cruce multipunto. Sin embargo, existen estudios que desaprueban esta técnica (DeJong & Spears, 1999), aunque se admite que el cruce de 2 puntos aporta una sustancial mejora con respecto al cruce de un solo punto, el hecho de añadir un mayor número de puntos de cruce reduce el rendimiento del algoritmo genético. El problema principal de añadir nuevos puntos de cruce radica en que es más fácil que los segmentos originados sean corrompibles, es decir, que por separado quizás pierdan las características de bondad que poseían conjuntamente. Hay que evitar, por lo tanto, romper dichos segmentos, denominados bloques constructivos.

Sin embargo, no todo son desventajas y añadiendo más puntos de cruce se consigue que el espacio de búsqueda del problema sea explorado con más intensidad.

* 1. *Cruce uniforme*

Es una técnica completamente diferente, cada gen de la descendencia tiene las mismas probabilidades de pertenecer a uno u otro padre.



*Figura 6: Cruce uniforme*

Aunque se puede implementar de muy diversas formas, la técnica implica la generación de una nueva mascara de cruce con valores binarios. Si en una de las posiciones de la máscara hay un 1, el gen situado en esa posición en uno de los genes descendientes se copia del primer padre; si por el contrario hay un 0 en el gen se copia del segundo padre. Para producir el segundo descendiente se intercambian los papeles de los padres, o bien se intercambia la interpretación de los unos y los ceros de la máscara de cruce.

La máscara de cruce puede no permanecer fija durante todo el proceso evolutivo, se genera de manera aleatoria para cada cruce. Se suele referir a ese tipo de cruce con las siglas UPX (Uniform Point Crossover).

* 1. *Cruces específicos de codificaciones no binarias*

Si se emplean genotipos compuestos por valores enteros o reales pueden definirse otro tipo de operadores de cruce como:

* *Media:* El gen de la descendencia toma el valor medio de los genes de los padres. Tiene la desventaja de que únicamente se genera un descendiente en el cruce de dos padres.
* *Media geométrica:* Cada gen de la descendencia toma como valor la raíz cuadrada del producto de los genes de los padres. Presenta el problema añadido de que signo dar al resultado si los padres tienen signos diferentes.
* *Extensión:* Se toma la diferencia existente entre los genes situados en las mismas posiciones de los padres y se suma al valor más alto o se resta del valor más bajo. Solventa el problema de generar un único descendiente.

1. *Algoritmos de Reemplazo*

Cuando en vez de trabajar con una población temporal se hace con una única población, sobre la que se realizan las selecciones e inserciones, deberá tenerse en cuenta que para insertar un nuevo individuo deberá de eliminarse previamente otro de la población. Existen diferentes métodos de reemplazo:

* *Aleatorio:* El nuevo individuo se inserta en un lugar escogido de manera aleatoria en la población.
* *Reemplazo de padres:* Se obtiene espacio para la nueva descendencia liberando el espacio ocupado por los padres.
* *Reemplazo de similares:* Una vez obtenido el ajuste de la descendencia se selecciona un grupo de individuos (entre seis y diez) de la población con un ajuste similar. Se reemplazan aleatoriamente los que sean necesarios.
* *Reemplazo de los peores:* De entre un porcentaje de los peores individuos de la población se seleccionan aleatoriamente los necesarios para dejar sitio a la descendencia.

1. *Copia*

La copia es una estrategia reproductiva para la obtención de una nueva generación a partir de la anterior. El porcentaje de copias de una generación a la siguiente es relativamente reducido, pues en caso contrario se corre el riesgo de una convergencia prematura de la población hacia ese individuo. De esta manera el tamaño efectivo de la población se reducirá notablemente y la búsqueda en el espacio del problema se focalizara en el entorno de ese individuo.

Lo que generalmente se suele hacer es seleccionar dos individuos para el cruce y, si este finalmente no tiene lugar, se insertan en la siguiente generación los individuos seleccionados.

1. *Elitismo*

El elitismo es un caso particular del operador de copia, consiste en copiar siempre el mejor o en su caso mejores individuos de una generación en la generación siguiente. De esta manera se garantiza que el proceso de búsqueda nunca dará un paso atrás en cuanto a la calidad de la mejor solución obtenida, sino que un cambio en esta siempre implicara una mejora.

Una variación de este proceso consiste en copiar al mejor de una generación en la siguiente, únicamente cuando tras el paso de una generación no se haya mejorado con los operadores de cruce o mutación la mejor solución de la generación actual.

1. *Mutación*

La mutación de uno individuo provoca que alguno de sus genes, generalmente uno solo, varié su valor de forma aleatoria.

Aunque se pueden seleccionar los individuos directamente de la población actual y mutarlos antes de introducirlos en la nueva población, la mutación se suele utilizar de manera conjunta con el operador de cruce. Inicialmente se seleccionan dos individuos de la población a realizar el cruce, si el cruce tiene éxito entonces uno de los descendientes, o ambos, se mutan con cierta probabilidad *Pm.* Se imita de esta manera el comportamiento que se da en la naturaleza, pues cuando se genera la descendencia siempre se produce algún tipo de error, por lo general sin mayor trascendencia, en el paso de la carga genética de padres a hijos.

La probabilidad de mutación es muy baja, generalmente menor al 1%, lo cual se debe sobre todo a que los individuos suelen tener un ajuste mejor después de mutados. Sin embargo, se realizan mutaciones para garantizar que ningún punto del espacio de búsqueda tenga una probabilidad nula de ser examinado.

La mutación más usual es el reemplazo aleatorio que consiste en variar aleatoriamente un gen de un cromosoma, si se trabaja con codificaciones binarias consistirá simplemente en negar un bit.

También es posible realizar la mutación intercambiando los valores de dos alelos del cromosoma. Con otro tipo de codificaciones no binarias existen otras opciones como:

* Incrementar o reducir a un gen una pequeña cantidad generada aleatoriamente.
* Multiplicar un gen por un valor aleatorio próximo a 1.

Aunque no es lo más común, existen implementaciones de algoritmos genéticos en las que no todos los individuos tienen los cromosomas de la misma longitud, lo que implica que no todos ellos codifican el mismo conjunto de variables. En este caso existen mutaciones adicionales como puede ser añadido un nuevo gen o eliminar uno ya existente.

1. *Evaluación*

Para el correcto funcionamiento de un algoritmo genético se debe poseer un método que indique si los individuos de la población representan o no buenas soluciones al problema planteado. Por lo tanto, para cada tipo de problema que se desee resolver deberá derivarse un nuevo método, al igual que ocurrirá con la propia codificación de los individuos.

La función de evaluación establece una medida numérica de la bondad de una solución, la cual recibe el nombre de ajuste, en la naturaleza el ajuste (o adecuación) de un individuo puede considerarse como la probabilidad de que ese individuo sobreviva hasta la edad de reproducción y se reproduzca, donde la probabilidad deberá estar ponderada con el número de individuos de la población genética.

En el mundo de los algoritmos genéticos se empleara esta medición para controlar la aplicación de los operadores genéticos, es decir, permitirá controlar el número de selecciones, cruces, copias y mutaciones llevadas a cabo.

La aproximación más común consiste en crear explícitamente una medida de ajuste para cada individuo de la población. A cada uno de los individuos se le asigna un valor de ajuste escalar por medio de un procedimiento de evaluación bien definido. También puede calcularse el ajuste mediante una manera “co-evolutiva”.

Se pueden diferenciar cuatro tipos de ajuste o fitness tales como:

* **Fitness Puro:***r()*

En la medida de ajuste establecida en la terminología natural del propio problema. Por ejemplo, supóngase una población de hormigas que deben llenar la despensa de cara al invierno, la bondad de cada hormiga será el número de piezas de comida llevadas por ella hasta el hormiguero.

En los problemas de maximización, como sería el de las hormigas mencionado anteriormente, los individuos con un fitness puro elevado serán los más interesantes. Al contrario, en los problemas de minimización interesaran los individuos con un fitness puro reducido.

* **Fitness Estandarizado:**

Para solucionar esta dualidad ante problemas de minimización p maximización se modifica el ajuste puro de acuerdo a la siguiente formula:

En el caso de problemas de minimización se emplea directamente la medida de fitness puro. Si el problema es de maximizar se resta una cota superior del error el fitness puro.

Empleando esta métrica la bondad de un individuo será menor cuanto más cerca este a cero el valor de ajuste. Por tanto, dentro de la generación , un individuo siempre será mejor que uno si se verifica que .

* **Fitness Ajustado:**

Se obtiene aplicando la siguiente transformación al fitness estandarizado:

De esta manera, el fitness ajustado tomará siempre valores del intervalo (0,1]. Cuando más se aproxime el fitness ajustado de un individuo a 1 mayor será su bondad.

* **Fitness Normalizado:**

Los diferentes tipos de fitness vistos hasta ahora hacen referencia únicamente a la bondad del individuo en cuestión. El fitness normalizado introduce un nuevo aspecto: indica la bondad de una solución con respecto al resto de soluciones representadas en la población. Se obtiene de la siguiente forma (considerando una población de tamaño M):

Al igual que el fitness ajustado, siempre tomara valores del intervalo [0,1], con mejores individuos cuanto más próximo este a la unidad. Pero a diferencia de antes, un valor cercano a 1 no solo indica que ese individuo represente una buena solución destacadamente mejor que las proporcionadas por el resto de la población. La suma de los valores de fitness normalizado de una población da siempre 1.

Este tipo de ajuste es empleado en la mayoría de los métodos de selección proporcionales al fitness.

**BUSQUEDA TABÚ**

Según (Díaz Fernández et al., 1996) la Búsqueda Tabú es un procedimiento meta-heurístico utilizado para guiar un algoritmo heurístico de búsqueda local para explorar el espacio de soluciones más allá de la optimalidad local. Fue propuesta inicialmente por (Glover, 1986) y desde entonces ha sido ampliamente usada en la solución de problemas de optimización combinatoria y de gran escala.

El énfasis en la exploración responsiva considerada en la búsqueda tabú deriva de la suposición de que una mala elección estratégica puede proporcionar más información que una buena elección al azar, dado que una elección estratégica puede proporcionar pistas útiles sobre como guiar la búsqueda hacia zonas prometedoras. Por lo tanto, la exploración responsiva integra los principios básicos de la búsqueda inteligente; explota las características de las soluciones buenas a la vez que explora nuevas regiones prometedoras.

Para lograr escapar de óptimos locales TS (Tabu Search) por sus siglas en ingles, se basa en metodologías de memorias flexibles. Esta memoria se puede clasificar en memoria a corto plazo y memoria a largo plazo, la memoria a corto plazo guarda en una lista tabú los últimos movimientos realizados de tal manera que se vetan por un número determinado de iteraciones, esto tiene como objetivo llevar un registro de los movimientos que pueden hacer que la solución regrese a un óptimo local de una iteración anterior y la memoria a largo plazo que lo que hace es diversificar la búsqueda hacia zonas que no han sido exploradas todavía, nuevamente buscando escapar de óptimos locales. Adicionalmente TS cuenta con un criterio de aspiración el cual permite levantar el veto a un movimiento antes de que se cumpla el número de iteraciones definidas en la lista tabú si realizar este movimiento genera una mejor solución.

A continuación se presentan los pasos básicos en el desarrollo de un algoritmo de búsqueda Tabú:

1. Selección de una solución inicial. Esta solución debe ser una solución factible.
2. Elección del entorno y generación de una nueva solución.
3. Evaluación de la función objetivo.
4. Actualización de la mejor solución y de la mejor de las mejores soluciones.
5. Criterio de finalización.
6. **Parámetros de entrada**
   1. Método para encontrar la solución inicial. El usuario puede decidir si la solución inicial se encuentra de forma aleatoria o usando la heurística constructiva de vecino más cercano.
   2. Tamaño de la lista Tabú. El usuario puede decidir entre una lista tabú dependiente la cual varia directamente proporcional al número de nodos (se define de tamaño n/2 donde n es el número de nodos de la instancia) o fija que será un numero entero definido por el usuario. Este tamaño de lista se tiene en cuenta a la hora de hacer el intercambio de nodos 2-opt para mejorar el balance.
   3. Número de iteraciones. Es el único criterio de parada de la meta-heurística el cual es un número de iteraciones definidas por el usuario.
7. **Determinar el tamaño de flota de vehículo**

Para cada instancia se calcula el número de vehículos necesarios de acuerdo con la capacidad de cada vehículo y a la demanda total de los clientes.

1. **Encontrar una solución inicial**

Para el desarrollo de este punto el usuario del programa puede decidir si la solución inicial se encuentra de forma aleatoria o usando la heurística de vecino más cercano.

**ESTRUCTURA DE LA BÚSQUEDA TABU**

1. **Uso de memoria**

Las estructuras de memoria de la búsqueda tabú funcionan mediante referencia a cuatro dimensiones principales, consistentes en la propiedad de ser reciente, en frecuencia, en calidad y en influencia. Las memorias basadas en lo reciente y en frecuencia se complementan la una a la otra para lograr el balance entre intensificación y diversificación que todo proceso de búsqueda heurística debe poseer. La dimensión de calidad hace referencia a la habilidad para diferenciar la bondad de las soluciones visitadas a lo largo del proceso de búsqueda, de tal forma, la memoria puede ser utilizada para la identificación de elementos comunes a soluciones buenas o a ciertos caminos que conducen a ellas. La calidad constituye un fundamento para el aprendizaje basado en incentivos, donde se refuerzan las acciones que conducen a buenas soluciones y se penalizan aquellas que, por el contrario, conducen a soluciones pobres. La flexibilidad de las estructuras de memoria mencionadas hasta el momento permite guiar la búsqueda en un entorno multi-objetivo, dado que se determina la bondad de una dirección de búsqueda particular mediante más de una función.

Por último, la cuarta dimensión de memoria, referida a la influencia, considera el impacto de las decisiones tomadas durante la búsqueda, no solo en lo referente a la calidad de las soluciones, sino también en lo referente a la estructura de las mismas. Este último uso de memoria es una característica importante de la búsqueda tabú que con frecuencia se olvida, pero que debería ser considerada incluso en los diseños más simples.

El uso de memoria en la búsqueda tabú es tanto explicita como implícita. En el primer caso, se almacenan en memoria soluciones completas, generalmente soluciones elite visitadas durante la búsqueda, mientras que en el segundo caso, se almacena información sobre determinados atributos de las soluciones que cambian al pasar de una solución a otra. Aunque en algunos casos, la memoria explicita es usada para evitar visitar soluciones más de una vez, esta aplicación es limitada dado que es necesario la implementación de estructuras de memoria muy eficientes para evitar requerimientos de memoria excesivos. Estos dos tipos de memoria son complementarios, puesto que la memoria explicita permite expandir los entornos de búsqueda usados durante un proceso de búsqueda local mediante la inclusión de soluciones elite, mientras que la memoria basada en atributos los reduce prohibiendo determinados movimientos.

1. **Intensificación y Diversificación**

Las estrategias de intensificación y diversificación constituyen dos elementos altamente importantes en un proceso de búsqueda tabú. Las estrategias de *intensificación* se basan en la modificación de reglas de selección para favorecer la elección de buenas combinaciones de movimientos y características de soluciones encontradas. Esto implica que es necesario identificar un conjunto de soluciones elite cuyos buenos atributos puedan ser incorporados a nuevas soluciones creadas, la permanencia al conjunto de soluciones elite se determina generalmente atendiendo a los valores de la función objetivo comparado con la mejor solución obtenida hasta el momento.

Por otro lado, las estrategias de *diversificación* tratan de conducir la búsqueda a zonas del espacio de soluciones no visitadas anteriormente y generar nuevas soluciones que difieran significativamente de las ya evaluadas.

Las estrategias de intensificación y diversificación interactúan para proporcionar puntos de apoyo fundamentales de memoria de largo plazo en búsqueda tabú.

1. **Fundamentos de búsqueda Tabú: Memoria a corto plazo**

Antes de comenzar a detallar los fundamentos de la búsqueda tabú, es necesario disponer de algunas definiciones y convenciones básicas.

* 1. *Memoria y clasificación Tabú*

Un proceso de búsqueda local basado únicamente en estrategias a corto plazo puede permitir que una solución sea visitada más de una vez, pero es probable que el entorno reducido sea diferente en cada una de las exploraciones. Cuando la memoria a corto plazo va acompañada de memoria a largo plazo, se reduce en gran medida la probabilidad de tomar decisiones que visiten repetidamente solo un subconjunto limitado del espacio de soluciones.

La función de memoria a corto plazo constituye uno de los elementos más importantes de la metodología de búsqueda tabú. Estas funciones aportan a la búsqueda la oportunidad de continuar más allá de la optimalidad local permitiendo la ejecución de movimientos de no mejora ligados a la modificación de la estructura de entorno de las siguientes soluciones.

* 1. *Memoria basada en lo reciente*

La memoria a corto plazo más utilizada generalmente en la literatura almacena los atributos de las soluciones que han cambiado en el proceso reciente. Este tipo de memoria a corto plazo se denomina *memoria basada en lo reciente*, la forma más habitual de explotar este tipo de memoria es etiquetando los atributos seleccionados de soluciones visitadas recientemente como *tabú-activos.*

Se considera que un atributo es *tabú-activo* cuando su atributo inverso asociado ha ocurrido dentro de un intervalo estipulado de lo reciente; un atributo que no es *tabú-activo* se llama *tabú-inactivo.* Aquellas soluciones que contengan atributos *tabú-activos,* o combinaciones particulares de los mismos, se convierten en soluciones tabú o prohibidas.

* 1. *Periodo tabú*

El periodo tabú puede ser diferente para diferentes tipos o combinaciones de atributos, y con un mayor nivel de desarrollo, pueden variar también sobre diferentes estados del proceso de búsqueda. Estas variaciones del período tabú de los atributos hace posible crear diferentes formas de balance entre las estrategias de memoria a corto y a largo plazo.

Independientemente de la estructura de datos usada, la cuestión clave para crear el estado tabú usando memoria basada en lo reciente es determinar un buen valor de o período tabú. Es posible reconocer que un periodo tabú es muy pequeño para una clase de problemas cuando se detectan repetitivos valores de la función objetivo, lo cual sugiere la aparición de ciclado en el proceso de búsqueda. De la misma forma, se detecta que un período tabú es muy grande cuando se produce un deterioro de la calidad de las soluciones encontradas, es posible establecer un rango de períodos intermedios para obtener un buen comportamiento de la búsqueda. Una vez obtenidos este rango de períodos tabú, un modo de proceder es seleccionar diferentes valores del rango en iteraciones diferentes.

Los elementos de la memoria a corto plazo, combinados con consideraciones de memoria a largo plazo, hacen de la búsqueda tabú un método con gran poder. Sin embargo, el enfoque inicial de memoria a corto plazo por sí mismo es capaz de generar soluciones de alta calidad.

Las reglas para determinar el período tabú , se clasifican en estáticas y dinámicas. Las reglas estáticas eligen un valor para que se mantiene fijo a lo largo de la búsqueda y las reglas dinámicas permiten que el valor de varíe. La variación del período tabú durante el proceso de búsqueda proporciona un método efectivo para inducir un balance entre examinar una región en detalle y mover la búsqueda hacia regiones del espacio diferentes.

* 1. *Criterios de aspiración*

Otro de los elementos fundamentales que permite al método de búsqueda tabú alcanzar sus mejores niveles de ejecución es la introducción de los criterios de aspiración durante el proceso de búsqueda. Los criterios de aspiración de introducen para determinar cuándo se pueden reemplazar los restricciones tabú, eliminando así la clasificación tabú aplicada a un movimiento, aunque gran parte de las aplicaciones que se encuentran en la literatura emplean únicamente un tipo simple de criterio de aspiración, que consiste en eliminar una clasificación tabú de un movimiento de ensayo cuando el movimiento conduce a una solución mejor que la mejor obtenida hasta ahora, hay otros criterios de aspiración efectivos para mejorar la búsqueda.

Una base para uno de estos criterios surge al introducir el concepto de *influencia*, que mide al grado de cambio inducido en la estructura de la solución o en la factibilidad. Un movimiento de alta influencia, que cambia significativamente la estructura de la solución actual, se ejemplifica mediante un movimiento que transfiera un objeto muy pesado de una caja a otra, tal movimiento puede no mejorar la solución actual, siendo menos probable conducir a una mejora cuando la solución actual sea relativamente buena. Se realizan movimientos de baja influencia mientras existan posibilidades de mejora significantes.

En el momento en el que se carezca de movimientos de mejora, los criterios de aspiración cambian para dar un mayor peso a los movimientos influyentes. Además, una vez que se ha realizado un movimiento influyente, cabe pensar que nos hemos desplazado a una región diferente del espacio de búsqueda y por lo tanto, deberían eliminarse las restricciones tabú establecidas previamente para movimientos menos influyentes.

Distinguimos entre *aspiraciones de movimiento y aspiraciones de atributo,* cuando se satisface una aspiración de movimiento, se revoca la clasificación tabú del movimiento. De la misma forma, cuando se satisface una aspiración de atributo, se revoca el estado tabú-activo del atributo.

***Criterios de aspiración ilustrativos***

* **Aspiración por defecto:**

Si todos los movimientos disponibles están clasificados tabú y no se han hecho admisibles mediante algunos otros criterios de aspiración, entonces se selecciona el movimiento “menos tabú”.

* **Aspiración por objeto:**

*Forma global:* se satisface una aspiración de movimiento, permitiendo que sea un candidato para la selección, si .

*Forma regional:* subdividimos el espacio de búsqueda en regiones , identificadas mediante cotas sobre los valores de funciones (o por intervalos de tiempos de búsqueda). Denotemos por el mínimo encontrado en , entonces , se satisface una aspiración de movimiento (para moverse hacia ) si .

* **Aspiración por dirección de búsqueda:**

Sea si el movimiento más reciente conteniendo a ē fue un movimiento de mejora y en otro caso. Se satisface una aspiración de atributo para (haciendo a tabú-inactivo) si y el movimiento ensayo actual es movimiento de mejora, es decir, si .

* **Aspiración por influencia:**

Sea o 1 según si el movimiento que establece el valor de es un movimiento de baja influencia o un movimiento de alta influencia. Además, sea , para o 1, igual a la iteración más reciente en la que fue realizado un movimiento de nivel de influencia , entonces una aspiración de atributo para se satisface si y . Para múltiples niveles de influencia la aspiración para se satisface si hay un tal que .

Las aspiraciones por Dirección de búsqueda y por influencia proporcionan aspiraciones de atributos en vez de aspiraciones de movimientos. En la mayoría de casos, las aspiraciones de atributos y movimientos son equivalentes. Sin embargo, se emplean diversos medios para probar estas dos clases de aspiraciones.

* + 1. **Refinamiento de los criterios de aspiración**

Algunas mejoras de los criterios ilustrados anteriormente proporcionan una oportunidad para realzar la potencia de la búsqueda tabú para aplicaciones que son más complejas o que ofrecen una recompensa por soluciones de muy alta calidad.

La creación de un estado tabú que varíe por grados, más que simplemente señalar un atributo para ser tabú-activo o tabú-inactivo, conduce a un refinamiento adicional de aspiración por dirección de búsqueda y aspiración por influencia. El estado tabú graduado está implícito en las variantes probabilísticas de la búsqueda tabú, donde el estado se expresa como una función de cómo un atributo se ha convertido recientemente o frecuentemente en tabú-activo y tabú-inactivo. Sin embargo, para emplear esta idea de realzar los criterios de aspiración precedentes, creamos un único estado tabú intermedio que cae entre los dos estados de tabú-activo y tabú-inactivo, en particular, cuando se satisface una aspiración para un atributo que en otro caso es tabú-activo se denomina atributo tabú pendiente.

Un movimiento que sería clasificado tabú si sus atributos tabú pendientes fueran tratados como tabú-activos, pero que no sería clasificado tabú en otro caso, es llamado movimiento tabú pendiente que puede ser tratado en uno de dos modos. En el enfoque menos restrictivo, tal movimiento no se previene de ser seleccionado, pero su estado cambia de tal manera que solo es candidato para selección si no existen movimientos de mejora excepto aquellos que son tabú. Un movimiento tabú pendiente debe ser, además, un movimiento de mejora para ser calificado para selección.

***Aspiración por admisibilidad fuerte:***

Las nociones precedentes conducen a un tipo adicional de aspiración. Definimos un movimiento como fuertemente admisible si:

1. Es admisible para ser seleccionado y no confía en criterios de aspiración para calificar para admisibilidad.
2. Califica para admisibilidad basado en aspiración global por objetivo, satisfaciendo .

Sea igual a la iteración más reciente en la que fue realizado u movimiento de no mejora y sea igual a la iteración más reciente en la que fue realizado un movimiento fuertemente admisible. Entonces, si , re-clasificamos cada movimiento tabú de no mejora como un movimiento tabú pendiente (permitiendo por tanto que sea un candidato para selección si no existe otro movimiento de mejora).

La desigualdad de la condición de aspiración precedente implica, que se ha realizado un movimiento de mejora fuertemente admisible desde el último movimiento de no mejora y por otro lado, que actualmente la búsqueda está generando una secuencia de mejora.

Este tipo de aspiraciones asegura que el método siempre procederá a un óptimo local siempre que se cree una secuencia de mejora que contenga al menos un movimiento fuertemente admisible.

* + 1. **Consideraciones especiales para la aspiración por influencia**

El criterio de aspiración por influencia puede ser modificado para crear un impacto considerable sobre su efectividad para ciertos tipos de aplicaciones, lo cual, deriva de la observación de que un movimiento característicamente es influyente en virtud de contener uno o más atributos influyentes. Bajo dichas condiciones, es apropiado considerar niveles de influencia definidos sobre los atributos, expresado por . Sin embargo, un movimiento puede derivar su influencia de la combinación única de los atributos involucrados y la aspiración por influencia preferiblemente transforma una aspiración de movimientos en vez de una aspiración de atributo.

En muchas aplicaciones, la influencia depende de una forma de conectividad, haciendo a sus efectos ser expresados principalmente sobre un rango particular, el cual se denomina rango *esfera de influencia* del movimiento o atributo asociado, bajo tales circunstancias, la aspiración por influencia debería estar limitada a modificar el estado tabú de atributos o la clasificación tabú de los movimientos que caen dentro de una esfera de influencia asociada.

El cambio del estado tabú continua dependiendo de las condiciones conocidas previamente, la influencia del atributo (o movimiento que lo contenga) debe ser menor que la de un movimiento anterior y la iteración para el atributo deber preceder a la iteración sobre la cual ocurrió el movimiento influyente anterior. Estas condiciones pueden ser registradas colocando un indicador para cuando se ejecuta el movimiento influyente, sin tener que comprobar otra vez para ver si es afectado por tal movimiento, cuando a se le reasigna un nuevo valor, el indicador es eliminado.

Son importantes las medidas de la influencia del movimiento y las caracterizaciones asociadas de esferas de influencia. Además, la influencia puede ser expresada como una función de los componentes de la memoria de la búsqueda tabú, como cuando un movimiento que contiene atributos que no han sido no frecuentemente ni recientemente tabú-activos puede ser clasificado como más altamente influyente, lo cual, fomenta una definición dinámica de la influencia, la cual varía según el estado actual de la búsqueda.

1. **Fundamentos de la búsqueda tabú: memoria a largo plazo**

En algunas aplicaciones, los componentes de la memoria a corto plazo son suficientes para producir soluciones de alta calidad. Sin embargo, la inclusión de la memoria a largo plazo, así como de las estrategias asociadas a la misma hace de la búsqueda tabú una estrategia más fuerte. En las estrategias de memoria a largo plazo, los entornos modificados de las soluciones actuales pueden contener soluciones que no estén en el entorno original, generalmente, se incluyen soluciones élite encontradas durante el proceso de búsqueda.

* 1. **Memoria basada en frecuencia**

Proporciona un tipo de información que complementa la información proporcionada por la memoria basada en lo reciente, ampliando la base para seleccionar movimientos preferidos, al igual que sucede con la memoria basada en lo reciente, la frecuencia a menudo esta ponderada o descompuesta en subclases teniendo en cuenta las dimensiones de calidad de la solución e influencia del movimiento.

Concebimos medidas de frecuencia como proporciones, cuyos numeradores representan contadores del número de ocurrencias de un evento en particular y cuyos denominadores generalmente representan uno de cuatro tipos de soluciones:

1. El número total de ocurrencias de todos los eventos representados por los numeradores (tal como el número de iteraciones asociadas).
2. La suma de los numeradores.
3. El máximo valor del numerador.
4. La medida del valor del numerador.

Los denominadores 3 y 4 dan lugar a lo que se puede llamar como frecuencias relativas, en los casos en los que los numeradores representan cuentas ponderadas, algunas de las cuales pueden ser negativas, los denominadores 3 y 4 se expresan como valores absolutos y el denominador 2 se expresa como la suma de los valores absolutos.

Denotemos por la secuencia de soluciones generadas en el momento presente del proceso de búsqueda y como una subsecuencia de esta secuencia de soluciones. Se toma la libertad de tratar a como un conjunto además de como una secuencia ordenada, los elementos de no son necesariamente elementos consecutivos de la secuencia de soluciones completa.

A modo de notación, denotemos por ) el conjunto de soluciones en para las cuales , y denotemos por la cardinalidad de este conjunto (el número de veces que recibe el valor sobre ). Análogamente, denotemos por el conjunto de soluciones en que resultan por un movimiento que cambia a . Finalmente, denotemos por y los conjuntos de soluciones en que contienen respectivamente como un o como En general, si representa cualquier atributo de una solución que puede tomar el papel de un o un para un movimiento y si representa un atributo de movimientos arbitrario denotado por (), entonces:

La cantidad constituye una medida de residencia, dado que identifica el número de veces que el atributo reside en las soluciones de . Correspondientemente, llamamos a la frecuencia que resulta de dividir tal medida por uno de los denominadores de 1 a 4 una frecuencia de residencia. Para el numerador , los denominadores 1 y 2 corresponden ambos a , mientras que los denominadores 3 y 4 son dados respectivamente por y por

Las cantidades , y constituyan medidas de transición, dado que identifican el número de veces que cambia de y/o a valores especificados. Asimismo, las frecuencias basadas en tales medidas son llamadas *frecuencias de transición,* los dominadores para crear tales frecuencias de las medidas precedentes incluyen *,* el número total de veces que los cambios indicados ocurren sobre S para diferentes valores y/o , y cantidades asociadas.

Las frecuencias de resistencia y transición en ocasiones trasmiten información relacionada, sin embargo, aunque a veces son confundidas en la literatura, en general tienen implicaciones diferentes. Una distinción es que las medidas de residencia, en contraste con las medidas de transición, no se refieren a si un atributo de soluciones particular de un elemento en la secuencia es un o un , o incluso si es un atributo que cambia en movimiento de a o de a . Solo es relevante que el atributo puede ser un o un en algún movimiento futuro, tales medidas pueden conducir a diferentes tipos de implicaciones dependiendo de la elección de la subsecuencia de .

Una frecuencia de resistencia alta, puede indicar que por ejemplo, un atributo es altamente atractivo si es una subsecuencia de soluciones de alta calidad, o puede indicar lo contrario si es una subsecuencia de soluciones de baja calidad. Por otro lado, una frecuencia de residencia que es alta (baja) cuando contiene tanto soluciones de alta como de baja calidad puede apuntar a atributo fortalecido (o excluido) que restringe al espacio de búsqueda y que necesita ser desechado (o incorporado) para permitir diversidad.

Desde el punto de vista de la simplificación del cómputo, cuando sea está formado por todas las soluciones generadas después de una iteración especificada, entonces puede mantenerse una medida de residencia actual y actualizada por referencia a valores del vector , sin la necesidad de incrementar un conjunto de contadores en cada iteración. Para un conjunto cuyas soluciones no vienen de iteraciones secuenciales, sin embargo, las medidas de residencia se calculan simplemente poniendo una etiqueta sobre los elementos de .

Una frecuencia de transición alta, en contraste con una frecuencia de residencia alta, puede indicar que un atributo asociado es un “llenador excelente”, que cambia dentro y fuera de la solución para ejecutar una función de buen ajuste. Tal atributo puede ser interpretado como el opuesto de un atributo influyente.

Los atributos que tienen mayores medidas de frecuencia, como aquellos que tienen mayores medidas de lo reciente (es decir, que ocurrieron en soluciones o movimientos más cercanos al presente), pueden iniciar un estado tabú-activo si está formado por soluciones consecutivas que finalizan con la solución actual. Sin embargo, la memoria basada en frecuencia típicamente encuentra su uso más productivo como parte de una estrategia de período más largo, la cual emplea incentivos además de restricciones para determinar que movimientos son seleccionados. En tal estrategia, las restricciones se convierten en penalizaciones de evaluación y los incentivos se convierten en mejoras de la evaluación, para alterar la base para calificar movimientos como atractivos o no atractivos.

La memoria basada en frecuencia por tanto es usualmente aplicada introduciendo estados tabú graduados, como un fundamento para definir valores de penalización e incentivos para modificar la evaluación de los movimientos. Existe una conexión natural entre este enfoque y el enfoque de memoria basada en lo reciente que crea estados tabú como una condición todo-o-ninguno. Si el período de un atributo en memoria basada en lo reciente está concebida como un umbral condicional para aplicar una penalización muy grande, entonces las clasificaciones tabú producidas por tal memoria pueden ser interpretadas con el resultado de una evaluación que se convierte fuertemente inferior cuando las penalizaciones están activadas. Es razonable anticipar que los umbrales condicionales deberían también ser relevantes para terminar los valores de penalizaciones y los incentivos en estrategias de período largo, la mayoría de las aplicaciones en el presente, sin embargo, usan un múltiplo lineal simple de una medida de frecuencia para crear un término de penalización o de incentivo.

* 1. **Estrategias de intensificación y diversificación simples**

Las funciones de intensificación y diversificación en la búsqueda tabú ya están implícitas en muchas de las prescripciones anteriores, pero se convierten especialmente relevantes en procesos de búsqueda de período largo. Las estrategias de intensificación crean soluciones agresivamente estimulando la incorporación de “atributos buenos”, en el período corto esto consiste en incorporar atributos que han recibido las mayores evaluaciones por enfoques y criterios descritos anteriormente, mientras que en el intermedio a largo período consiste en incorporar atributos de soluciones de subconjuntos élite seleccionados. Por otro lado, las estrategias de diversificación generan soluciones que incorporan composiciones de atributos significativamente diferentes a los encontrados previamente durante la búsqueda, estos tipos de estrategias se contrapesan y refuerzan mutuamente a varias formas.

Para una estrategia de diversificación elegimos como un subconjunto significativo de la secuencia de solución completa; por ejemplo, la secuencia entera empezando con el primer óptimo local, o la subsecuencia formada por todos los óptimos locales.(Para ciertas estrategias basadas en medidas de transición, puede estar formado por la subsecuencia que contiene cada sucesión intacta máxima de movimientos de no-mejora que inmediatamente siguen con óptimo local, concentrándose en para estos movimientos).

Para la estrategia de intensificación elegimos como un subconjunto pequeño de soluciones élite (óptimos locales de buena calidad) que comparten un gran número de atributos comunes, y en segundo lugar cuyos miembros pueden alcanzarse uno de otro mediante números de movimientos relativamente pequeños, independientes de si estas soluciones caen cerca la una de la otra en la secuencia de la solución. Por ejemplo, las colecciones de tales subconjuntos pueden ser generadas por procedimientos de agrupamiento, seguido del uso de un enfoque de procesamiento paralelo para tratar cada seleccionado por separado.

La medida de frecuencia puede extenderse para representar medidas combinadas de lo reciente y de lo frecuente. La memoria basada en lo reciente, almacenando datos de , puede también referirse a cambios que han ocurrido más lejos en el pasado además de aquellos que han ocurrido más recientemente, aunque estas medidas están ya implícitamente combinadas cuando se unen las penalizaciones y los incentivos basados en medidas de frecuencia con clasificaciones tabú basadas en medidas de lo reciente, como un fundamento para seleccionar movimientos actuales, es posible que otras formas de combinación sean superiores.

* 1. **Aspectos más avanzados de intensificación y diversificación**

Los métodos de intensificación y diversificación que utilizan penalizaciones e incentivos representan solo una clase de tales estrategias. Una colección mayor surge de la consideración directa de los objetivos de intensificación y diversificación. Es importante hacer una distinción entre diversificación y aleatorización.

*Diversificación frente a aleatorización:* cuando la búsqueda tabú busca una colección de soluciones diversas, es muy diferente de cuando busca una colección de soluciones aleatorias. En general, estamos interesados no solo en colecciones diversas sino en secuencias diversas, dado que frecuentemente el orden en el que se examinan los elementos es importante en TS. Esto ocurre, por ejemplo, cuando buscamos identificar una secuencia de nuevas soluciones de forma que cada solución sucesiva sea *maximalmente* diversa en relación a todas las soluciones previamente generadas, esto incluye posibles referencias a un conjunto base de soluciones previamente generadas.

El concepto de diversificación se aplica también a la generación de una secuencia diversa de números o a un conjunto diverso de puntos entre los vértices del hipercubo unidad.

*Refuerzo por restricción:* uno de los primeros tipos de estrategias de intensificación, caracterizada en términos de explotar variables fuertemente determinadas y consistentes, comienza seleccionando un conjunto como indicado para determinar una penalización y una función de incentivo, es decir, un conjunto formado por soluciones élite agrupadas a través de una medida de clasificación. En vez de crear penalizaciones e incentivos, con el objetivo de incorporar atributos a la solución actual que tenga altas medidas de frecuencia sobre , el método de refuerzo por restricción opera estrechando el rango de posibilidades permitidas añadiendo y quitando tales atributos. Por ejemplo si tiene una alta frecuencia sobre solo para un pequeño número de valores de , entonces los movimientos se restringen permitiendo a tomar solo uno de estos atributos en la definición de un . Por tanto, si es una variable 0-1 con una medida de frecuencia alta sobre para uno de sus valores, entonces este valor se hará fijo una vez que exista un movimiento admisible que permita que se asigne dicho valor. Otras asignaciones pueden permitirse por una variable de aspiración por defecto, si el conjunto actual de alternativas restringidas es inaceptable.

La consideración inicial sugiere que este método de restricción no ofrece nada más allá de las opciones disponibles por penalización e incentivos. No obstante, el método puede conseguir más que esto; primero, las restricciones explicitas pueden acelerar sustancialmente la ejecución de los pasos de elección reduciendo el número de alternativas examinadas. Segundo, muchos problemas se simplifican y colapsan una vez que se introducen un número de restricciones explicitas, permitiendo que las implicaciones estructurales salgan a la superficie, permitiendo que estos problemas se resuelvan más fácilmente.