**Resumen algoritmos genéticos**

**Introducción**

Desde que las computadoras fueron inventadas nos preguntamos qué son capaces de aprender. Los algoritmos genéticos parten de ésta búsqueda del aprendizaje de máquina y se basan en la evolución simulada.

La idea principal de estos algoritmos es, dada una población inicial, separar a los individuos mas aptos y combinar propiedades de los mismos creando así nuevos individuos que conserven, preferentemente, las mejores propiedades de la anterior generación.

Para determinar los individuos a seleccionar, se debe establecer una función de puntuación para los mismos (función fitness). Luego, para generar a los individuos de la nueva generación, se les realiza operaciones como mutaciones y cruzamientos aleatorios.

Los algoritmos genéticos han sido aplicados exitosamente en gran variedad de tareas de aprendizaje y en problemas de optimización.

**Analogía entre algoritmos genéticos y lo que pretenden simular**

El fenotipo de un individuo es el resultado de la interacción del medio ambiente en que se desarrolla y la herencia que recibe de sus ancestros. Dicho fenotipo está determinado por las proteínas que produce y está definido en la información genética de cada una de las células del individuo. A su vez, la información acerca de las proteínas que se producirán está contenida en los cromosomas. En cada célula existen dos juegos de cromosomas que definen las mismas características. Un cromosoma es una larga molécula de ADN formada por proteínas que definen los genes. El conjunto de todos los cromosomas, es decir, de toda la información genética de un individuo se llama genoma y el conjunto de genes contenidos en el genoma se denomina genotipo que es que determina en buena medida el fenotipo del individuo.

Es necesario codificar de alguna manera el dominio del problema para obtener estructuras manejables que puedan ser manipuladas por el algoritmo genético. Cada una de estas estructuras constituye el equivalente al genotipo de un individuo en términos biológicos. El elemento del dominio del problema al que se mapea este genotipo es el análogo al fenotipo. Es frecuente que el código de los elementos del dominio del problema utilice un alfabeto binario (0’s y 1’s).

Una vez que se ha definido la manera de codificar los elementos del dominio del problema y se conoce la forma de pasar de un elemento a su código y viceversa, es necesario fijar un punto de partida. Los algoritmos genéticos manipulan conjuntos de códigos (poblaciones de códigos) en generaciones sucesivas. El algoritmo se encargará de favorecer la aparición en la población de códigos que correspondan a elementos del dominio que estén próximos a resolver el problema. Es decir, dicho algoritmo recibirá como entrada una población de códigos y a partir de ésta generará nuevas poblaciones, donde algunos códigos desaparecerán mientras que otros, que se mapean en mejores soluciones posibles, aparecen con más frecuencia hasta que se encuentra una satisfactoria o hasta que se cumpla alguna otra condición de terminación. Los elementos de la población serán llamados individuos y a los códigos se les denominará indistintamente cromosomas, genotipo, genoma o código genético.

En la naturaleza hay individuos más hábiles que otros para sobrevivir y al igual que sucede en la naturaleza, en los algoritmos genéticos es necesario establecer algún criterio que permita decidir cuáles de las soluciones propuestas en una población son mejores respecto del resto de las propuestas y cuáles no lo son.

Para determinar cuáles de estos individuos corresponden a buenas propuestas de solución y cuáles no, es necesario calificarlos de alguna manera. Cada individuo de cada generación de un algoritmo genético recibe una calificación o una medida de su grado de adaptación (fitness). Éste es un número real no negativo que será más grande cuanto mejor sea la solución propuesta por dicho individuo.

Una vez calificados todos los individuos de una generación, el algoritmo debe seleccionar a los individuos más calificados para que tengan mayor oportunidad de reproducción. De esta forma se incrementa la probabilidad de tener individuos ‘buenos’ en el futuro. Si de una determinada generación se seleccionaran sólo aquellos con una calificación mayor o igual que cierto número *c* para pasarlos a la siguiente generación, es claro que en ésta la calificación superará *c* y por tanto al promedio de la generación anterior. La selección explota el conocimiento que se ha obtenido hasta el momento, procurando elegir lo mejor que se haya encontrado, elevando así el nivel de adaptación de toda la población.

La cruza de los códigos genéticos de individuos exitosos favorece la aparición de nuevos individuos que heredan de sus ancestros características deseables. En el contexto de los algoritmos genéticos reproducirse significa que, dados dos individuos seleccionados en función de su grado de adaptación, éstos pasen a formar parte de la siguiente generación o, al menos, mezclen sus códigos genéticos para generar ‘hijos’ que posean un código híbrido. Es decir, los códigos genéticos de los individuos se cruzan.

Para la réplica un cromosoma ya existente interviene una enzima denominada ADN-polimerasa. La molécula de ADN tiene forma de doble hélice y dicha enzima se encarga de abrir por la mitad la molécula y replicar el cromosoma. Ocasionalmente la ADN-polimerasa comete un error que puede ser causado por radiaciones energéticas externas o sustancias extrañas. La alteración de dicha molécula de ADN constituye una mutación que puede manifestarse en el fenotipo y hacer al individuo diferente del resto de sus congéneres. Por lo general las mutaciones son desfavorables e incluso letales para el organismo mutante pero a veces pueden no serlo y conferir a dicho organismo alguna ventaja que le permita sobrevivir más fácilmente en su medio. Dicha característica será transmitida a sus descendientes y se habrá producido un pequeño paso evolutivo.

Ocasionalmente algunos elementos del código de ciertos individuos de un algoritmo genético se alteran a propósito. Éstos se seleccionan aleatoriamente en lo que constituye el símil de una mutación. El objetivo es generar nuevos individuos que exploren regiones del dominio del problema que probablemente no se han visitado aún. Esta exploración no presupone conocimiento alguno. Aleatoriamente se buscan nuevas soluciones posibles que quizá superen las encontradas hasta el momento. Esta es una de las características que hacen aplicables los algoritmos genéticos a gran variedad de problemas: no presuponer conocimiento previo acerca del problema a resolver ni de su dominio, no sólo en la mutación sino en el proceso total. De hecho, el problema a resolver sólo determina la función de evaluación y la manera de codificar las soluciones posibles. El resto de los subprocesos que constituyen el algoritmo son independientes y universalmente aplicables.

**Definición y factores limitantes**

Un **algoritmo genético** consiste en una función matemática o una rutina de software que toma como entradas a los ejemplares y retorna como salidas cuales de ellos deben generar descendencia para la nueva generación.

Algunas versiones más complejas de algoritmos genéticos generan un ciclo iterativo que directamente toma a la especie (el total de los ejemplares) y crea una nueva generación que reemplaza a la antigua una cantidad de veces determinada por su propio diseño. Una de sus características principales es la de ir perfeccionando su propia heurística en el proceso de ejecución, por lo que no requiere largos períodos de entrenamiento especializado por parte del ser humano, principal defecto de otros métodos para solucionar problemas, como los Sistemas Expertos.

La aplicación más común de los algoritmos genéticos ha sido la solución de problemas de optimización, en donde han mostrado ser muy eficientes y confiables. Sin embargo, no todos los problemas pudieran ser apropiados para la técnica, y se recomienda en general tomar en cuenta las siguientes características del mismo antes de intentar usarla:

* Su espacio de búsqueda debe estar delimitado dentro de un cierto rango.
* Debe poderse definir una función fitness que nos indique cómo es de buena o mala es una cierta respuesta.
* Las soluciones deben codificarse de una forma que resulte relativamente fácil de implementar en la computadora.

El primer punto es muy importante, y lo más recomendable es intentar resolver problemas que tengan espacios de búsqueda discretos aunque éstos sean muy grandes. No sería conveniente la utilización de algoritmos genéticos en las siguientes situaciones:

* Si se requiere llegar forzosamente al máximo global en el caso de que estemos maximizando una función.
* Si conocemos la función de optimización.
* Si el problema está muy delimitado y se presta a un tratamiento analítico (funciones de una variable).
* Si la función es suave y convexa.
* Si el espacio es limitado, entonces es mejor enumerar todas las soluciones posibles.

**Principales áreas de aplicación**

1. **Análisis discriminante:**

El análisis discriminante forma parte del conjunto de técnicas estadísticas diseñadas para resolver el problema de clasificación. Se caracteriza por estudiar la relación entre una variable categórica dependiente (el grupo de clasificación) y un conjunto de variables reales (posiblemente vectoriales) independientes. Estas describen a cada uno de los individuos a clasificar. A través de esta técnica, se construyen reglas de decisión que permiten discriminar o separar grupos mediante funciones de las variables observadas**,** minimizando la probabilidad de clasificación errónea, o bien maximizándola si la clasificación es correcta. Las reglas más usuales se construyen a partir de un modelo probabilístico, y de muestras de entrenamiento.

1. **Problema del agente viajero:**

El problema del agente viajero, también denominado TSP (Travelling Salesman Problem), consiste en, dada una colección de ciudades, determinar el recorrido de coste mínimo, visitando cada ciudad exactamente una vez y volviendo al punto de partida.

A lo largo de los años el problema del agente viajero ha ocupado la mente de numerosos investigadores. Los motivos son varios. En primer lugar, el TSP es un problema muy sencillo de enunciar, pero muy difícil de resolver. En segundo lugar, el TSP es aplicable a una gran variedad de problemas de planificación. Finalmente, se ha convertido en una especie de problema test, es decir los nuevos métodos de optimización combinatoria son a menudo aplicados al TSP con objeto de tener una idea de sus potencialidades.

1. **Planificación de actividades:**

Los problemas de planificación se encuentran en una gran variedad de campos, incluyendo la manufactura y el servicio industrial. Los problemas de planificación son numerosos y variados. En términos más amplios, la planificación implica el reparto de recursos durante un período de tiempo para llevar a cabo un conjunto de actividades. En general, no hay un algoritmo general que garantice dar una solución óptima y hacerlo en un tiempo polinomial. Por tanto, los problemas de planificación son principales candidatos para la aplicación de la tecnología de la Inteligencia Artificial.

1. **Entrenamiento de redes neuronales profundas:**

Los algoritmos genéticos resultan ser una buena alternativa a otros algoritmos utilizados en esta área, los cuales son basados en métodos de gradientes, parecidos a diferencias finitas. Si bien podria esperarse que los AGs funcionen peor, por el hecho de ser tan simples y no estar basados en gradientes, resulta que son una alternativa altamente competitiva, y en muchos casos funcionan mejor que los metodos utilizados tradicionalmente en tareas de RL (reinforcement learning), los cuales estan basados en gradientes.

1. **Locomoción humanoide:**

Los AGs pudieron resolver el desafiante punto de referencia de control contínuo de locomoción humanoide. Si bien estos algoritmos produjeron robots que podian caminar bien, tomó aproximadamente 15 veces mas tiempo que utilizando algoritmos de ES (evolution strategies, otra rama de algoritmos evolutivos), y se obtuvo un rendimiento ligeramente peor que con los ultimos mensionados. Se requiere investigación futura para entender bien por qué.