Capítulo 8

Optimización

8.1. Algoritmos de optimización

En esta sección analizaremos una clase de algoritmos de optimización (o de búsqueda) agrupados bajo el término de *enfriamiento estocástico*.

8.1.1. Esquema frío

Consideremos un conjunto finito o infinito E de estados (o soluciones) y una función $C : E \to \mathbb{R}$ una función de costo (utilidad) a ser minimizada (maximizada). Buscamos algún valor $i^* \in E$ tal que

$$C(i^*) \leq C(i) \qquad \forall i \in E,$$

es decir la mejor solución. Cuando el espacio E es muy grande o infinito, es poco práctico imposible enumerar todos los valores $\{C(i)\}_{i \in E}$ para elegir al valor mínimo y al estado minimizante. Describimos a continuación un método simple para intentar encontrarlo.

Definimos un sistema de vecindades, que son una colección de conjuntos $\{N(i)\}_{i \in E}$ (donde $N(i) \subseteq E$), tales que

1. $i \notin N(i)$, es decir i no es vecino de sí mismo,
2. y si $i \in N(j)$ ocurre que $j \in N(i)$, es decir que el que i sea vecino de j implica que j sea vecino de i.

Pediremos además que el sistema de vecindades $\{N(i)\}_{i \in E}$ sea comunicante, es decir, que para todo i, j en E, existe una sucesión $i_1, ..., i_m$ en E tales que

$$i_1 \in N(i),\, i_2 \in N(i_1),\, ...,\, j \in N(i_m),$$

en otras palabras, de un estado i a otro j, se puede llegar a través de un número finito de vecinos.

El *esquema frío* consiste en el siguiente algoritmo.

Tenemos dos observaciones. La primera es la siguiente: en el caso en que tenemos un espacio E no numerable y elegimos como sistema de vecindades a $N(i) := E \setminus \{i\}$ (es decir todos los estados diferentes i, j son

Algoritmo 54 :Esquema frío

1: Iniciamos con un estado o solución i.
2: Elegimos un vecino $j \in N(i)$ aleatoriamente (no necesariamente de forma equiprobable).
3: Comparamos a $C(i)$ y $C(j)$ y elegimos a i o j como aquél que tiene el menor costo.
4: Repetimos el algoritmo desde el segundo paso

vecinos) y eligimos a un vecino de forma equiprobable, el Algoritmo 54 con n iteraciones, es equivalente a sortear $n-1$ estados o soluciones de forma independiente, añadir la solución inicial, y elegir entre esos estados a aquel que minimiza la función C.

La segunda observación nos habla de la gran limitación de este algoritmo. En el caso en que tengamos una solución i tal que

$$C(i) \leq C(j) \qquad \forall j \in N(i),$$

es decir un estado que es un mínimo local, el algoritmo se queda atrapado en ese valor, cuando posiblemente exista una mejor solución global.

8.1.2. Algoritmos evolutivos

Los algoritmos evolutivos se pueden conceptualizar dentro de los algoritmos de optimización o de búsqueda. Presentamos a continuación una versión simple, que es una generalización del esquema frío.

Algoritmo 55 Evolutivo

1: Iniciamos con un estado o solución i.
2: Con probabilidad α (muy baja) elegimos a j uniformemente sobre todo el espacio de soluciones E (*mutación*). Con probabilidad $1-\alpha$ (muy alta) elegimos a un vecino $j \in N(i)$ aleatoriamente (*descendencia*).
3: Comparamos a $C(i)$ y $C(j)$ y elegimos a i o j como aquél que tiene el menor costo.
4: Repetimos el algoritmo desde el segundo paso.

La interpretación es la siguiente: Teniendo una solución o individuo i podemos, o bien tener un descendiente (que debe tener características cercanas a i) con probabilidad muy alta, o bien encontramos a otro individuo proveniente de una mutación con probabilidad muy baja. Ambos compiten entre sí, y el que tiene mejor desempeño sobrevive a la siguiente generación.

La gran ventaja de este algoritmo sobre el esquema frío es que no nos quedamos atascados en mínimos locales, pero a diferencia de la búsqueda independiente, la descendencia nos ayuda a utilizar las iteraciones pasadas para un mejor desempeño del algoritmo.

8.1.3. Enfriamiento Estocástico

Este algoritmo generaliza el esquema frío y a su vez, es un caso particular del algoritmo del Metropolis-Hastings.

Dado un sistema de vecindades $\{N(i)\}_{i\in E}$ comunicante, consideremos una cadena de Markov definida sobre E con transición $\{q_{i,j}\}_{i,j\in E}$ tal que $q_{i,j} > 0$ si y sólo si $i = j$ o bien j está en $N(i)$. Definamos un valor $T > 0$ que representará en este caso la temperatura.

Definimos $P_{i,j}$ por

$$P_{i,j}(T) := 1_{E\setminus\{i\}}(j)q_{i,j}\alpha_{i,j}(T) + 1_{\{i\}}(j)\sum_{k\neq i} q_{i,k}\alpha_{i,k}(T)$$

donde

$$\alpha_{i,j}(T) := \text{mín}\Big(1, \frac{e^{\frac{C(i)}{T}}}{e^{\frac{C(j)}{T}}}\Big).$$

Desde la perspectiva de los algoritmos de Hastings, el parámetro de la temperatura T no influye, es decir la prueba utilizada ahí sigue funcionando en este caso.

Analicemos ahora a las probabilidades de aceptación $\alpha_{i,j}(T)$. En el caso en que $C(i) \geq C(j)$ ocurre que $\alpha_{i,j}(T) = 1$, por lo que aceptamos j (lo cuál es consistente con que queremos minimizar). En el caso en que $C(i) < C(j)$, a diferencia del algoritmo frío, no rechazamos directamente a j. Aceptamos j con probabilidad

$$e^{\frac{C(i)-C(j)}{T}},$$

aunque j sea peor solución que i.

La idea es que, con temperatura positiva, permitimos al sistema de búsqueda explorar regiones aún cuando las soluciones no estén mejorando de forma monónota en cada iteración. La maginitud del parámetro T nos dice que tan posible es para el algoritmo explorar regiones en donde es poco verosímil encontrar el mínimo deseado.

El método de enfriamiento estocástico consiste en ir *disminuyendo* lentamente la temperatura T, de tal forma que al principio exploramos muy diversas regiones, pero a medida en que las iteraciones avanzan buscamos cáda vez más localmente. El algoritmo es el siguiente:

Algoritmo 56 : Enfriamiento Estocástico

1: Iniciamos $X = i$, temperatura $T = t$, iteraciones realizadas $k = 0$.
2: Elegimos a un vecino $j \in N(i)$ con probabilidad $q_{i,j}$.
3: Con probabilidad $\alpha_{i,j}(T)$ hacemos $X = j$.
4: Hacemos $T = T - \frac{t}{n}$, $k = k + 1$.
5: Repetimos hasta que $k = n$.
6: Devolvemos X.

8.2. Algoritmo EM (Optimización en Estadística)

Un problema importante tanto en el enfoque clásico como bayesiano de la estadística es el de maximización, donde desde el punto de vista clásico se buscan estimadores máximo verosímiles y en Bayesiana se desea la maximizacón de la utilidad esperada en la toma de decisiones por ejemplo.