

Universidade Federal do Pará
Instituto de Tecnologia -ITEC
Disciplina: IA Bio-Inspirada e Otimização.

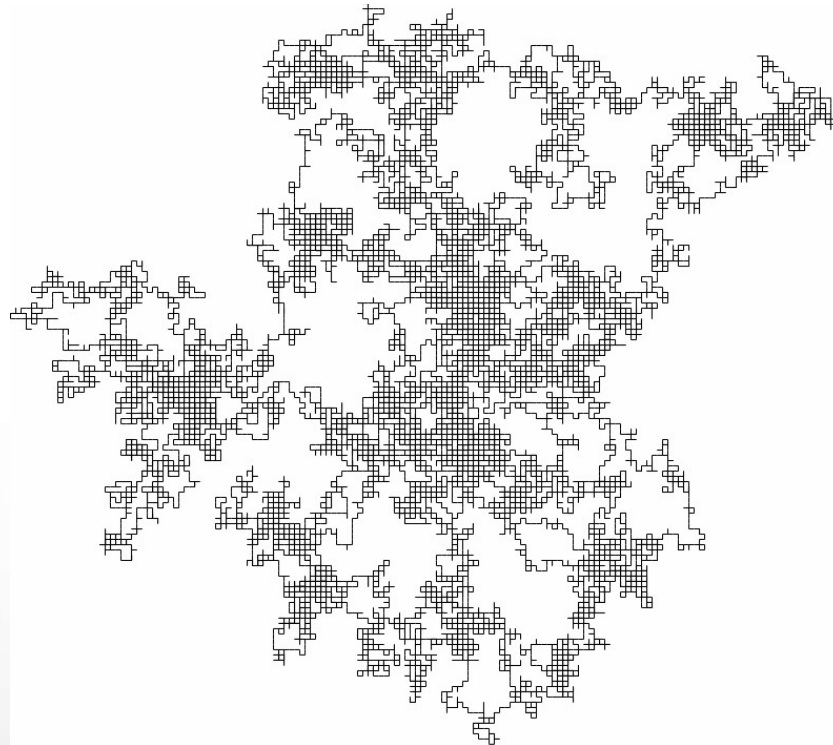


Random Walks e Lévy Flights

Profa: Jasmine Araújo.
Prof: Glauco.

Sumário

- Random Variables
- Random Walks
- Lévy Distribution and Lévy Flights
- Otimização com Cadeias de Markov




Otimização com Cadeias de Markov

- O random walk simples que discutimos na aula anterior pode ser considerado como uma cadeia de Markov.
- Uma variável aleatória ζ é um processo de Markov se a probabilidade de transição, do estado $\zeta_t = S_i$ no tempo t para outro estado $\zeta_{t+1} = S_j$ depende somente do estado atual ζ_t .

$$\begin{aligned} P(i, j) &\equiv P(\zeta_{t+1} = S_j | \zeta_0 = S_p, \dots, \zeta_t = S_i) \\ &= P(\zeta_{t+1} = S_j | \zeta_t = S_i), \end{aligned}$$

o qual é independente dos outros estados antes de t .

Otimização com Cadeias de Markov




- A sequência de variáveis aleatórias $\zeta_0, \zeta_1, \dots, \zeta_n$ geradas por um processo de Markov é chamada de cadeia de Markov.
- A probabilidade de transição $P(i, j) \equiv P(i \rightarrow j) = P_{ij}$, também é chamada de transição de kernel da cadeia de Markov.
- Se o random walk for reescrito como um movimento aleatório governado por w_t , o qual depende da probabilidade de transição P ,

Otimização com Cadeias de Markov



- $S_{t+1} = S_t + w_t$
- que de fato tem as propriedades da cadeia de Markov, logo o random walk é uma cadeia de Markov.
- Então, para resolver um problema de otimização nós precisamos procurar a solução realizando um random walk começando de um início bom mas uma solução de determinada de forma aleatória.

Otimização com Cadeias de Markov



- Entretanto random walk simples ou vedados não são eficientes.
- Para ser computacionalmente eficiente e efetivo em procurar por novas soluções, temos que manter as melhores soluções encontradas e aumentar a mobilidade do random walk para então explorar de forma eficiente o espaço de busca.
- Uma forma de controlar o passeio é necessária de uma maneira que ele se mova em direção a soluções ótimas mais rapidamente.

Otimização com Cadeias de Markov



- Esses são os desafios da maior parte dos algoritmos metaheurísticos.

Markov Chain Algorithm for Optimization

Start with $\zeta_0 \in S$, at $t = 0$

while *(criterion)*

Propose a new solution Y_{t+1} ;

Generate a random number $0 \leq P_t \leq 1$;

$$\zeta_{t+1} = \begin{cases} Y_{t+1} & \text{with probability } P_t \\ \zeta_t & \text{with probability } 1 - P_t \end{cases}$$

end

Simulated Annealing



- É uma das metaheurísticas mais novas e bem usadas.
- Ela é baseada em trajetória e busca aleatória para otimização global.
- Ele imita o processo de recozimento em processamento de material quando o metal esfria e congela em uma estrutura cristalina com a mínima energia e o maior tamanho de cristal para reduzir os defeitos em estruturas metálicas.
- O processo de recozimento envolve o cuidadoso controle da temperatura e sua taxa de resfriamento, frequentemente chamado de annealing schedule.

Simulated Annealing

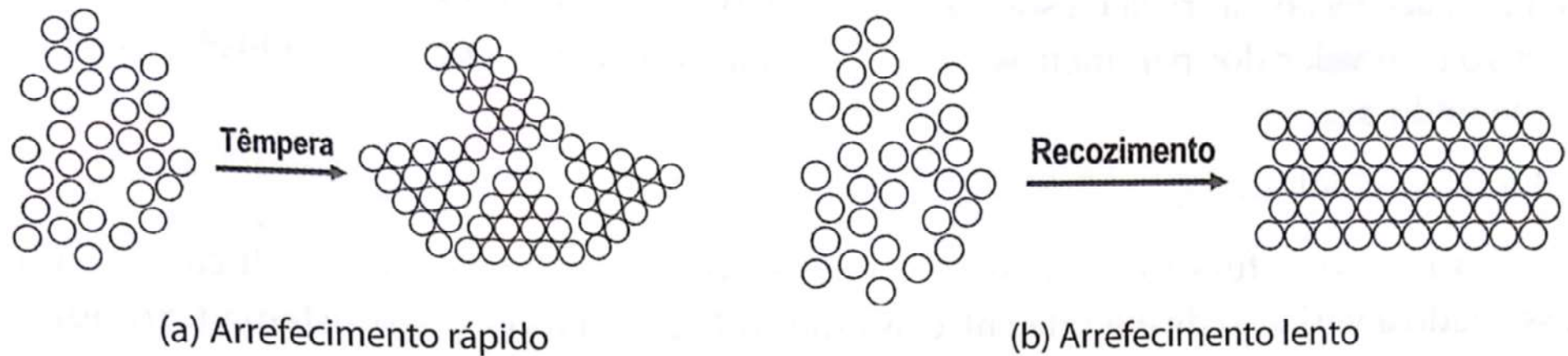


Figura 1: As configurações cristalinas.
Fonte: Goldberg *et al.* (2016, p. 107).



Simulated Annealing

- Annealing and Boltzmann Distribution
 - O primeiro desenvolvimento do simulated annealing por Kirkpatrick, Gelatt e Vecchi foi em 1983.
 - Desde então tem sido aplicado a todas as áreas de otimização.
 - local ótimo versus global ótimo.
 - Essencialmente, o simulated annealing é um algoritmo de busca via uma cadeia de Markov o qual converge sob determinadas condições.
 - isso equivale a lançar algumas bolas quicando sobre uma paisagem, à medida que as bolas quicam e perdem energia, elas se estabilizam em alguns mínimos locais.
 - Se as bolas saltarem vezes suficientes e perderem energia lentamente, algumas das bolas acabarão por cair nos locais globalmente mais baixos, portanto, os mínimos globais serão alcançados.



Simulated Annealing

- Annealing and Boltzmann Distribution
 - A ideia básica do SA é usar a busca aleatória em termos da cadeia de Markov, que não apenas aceita mudanças na melhoria da função objetivo mas também mantém mudanças que não são o ideal.
 - Em um problema de minimização, por exemplo, algum movimento ou mudança que decresce o valor da função objetivo f será aceito com uma probabilidade p .
 - Essa probabilidade p também chamada de probabilidade de transição é determinada por:
 - $$p = e^{-\frac{\Delta E}{k_B T}}$$
 - onde k_B é a constante de Boltzmann e por simplicidade pode-se usar somente k e o valor comumente usado é $k=1$.
 - T é a temperatura para controlar o processo de recozimento.



Simulated Annealing

- ΔE é mudança no nível de energia e é a probabilidade de transição baseado na distribuição de Boltzmann na mecânica estatística.
- A maneira mais simples de ligar ΔE com a mudança na função objetivo é:
- $\Delta E = \gamma \Delta f$ onde γ é um valor real constante.
- Por simplicidade também $\gamma=1$.
- Por esta razão, a probabilidade p é definida como:
$$p(\Delta f, T) = e^{-\Delta f/T}$$

Simulated Annealing



- se aceitamos ou não uma mudança, comumente se usa um valor aleatório r como limite. Por esta razão:
 - se $p = e^{-\Delta f/T} > r$ então o movimento é aceito.

Simulated Annealing



- Parâmetros:
 - A temperatura inicial T é muito importante.
 - Para uma dada mudança Δf , se T é muito alto ($T \rightarrow \infty$) então $p=1$
 - Isso indica que todas as mudanças serão aceitas
 - se T é muito baixo ($T \rightarrow 0$) então algum $\Delta f > 0$ será raramente aceita, pois $p \rightarrow 0$ e a diversidade da solução é limitada,
 - No entanto, alguma melhoria no Δf quase sempre será aceita.
 - No caso do ($T \rightarrow 0$) corresponde ao método baseado em gradiente, porque somente boas soluções são aceitas e o sistema é essencialmente subindo ou descendo uma colina.
 - Se o T é muito alto, o sistema está em um estado de energia muito alto e a mínima não é facilmente encontrada.

Simulated Annealing



- Parâmetros:
 - Se o T é muito baixo pode cair em um mínimo local, não necessariamente é um mínimo global.
 - E não existe energia suficiente para pular desse estado de mínimo local para explorar outros mínimos inclusive o global.
 - Então uma temperatura inicial apropriada deveria ser calculada.
- Outra importante tarefa é como controlar o processo de recozimento ou resfriamento de modo que o sistema esfrie gradualmente de uma temperatura mais alta para finalmente congelar em um estado mínimo global
- Existem muitas maneiras de controlar a taxa de resfriamento ou decrescer a temperatura.

Simulated Annealing



- Parâmetros:
 - Dois tipos de resfriamento são linear e geométrico:
 - Para um esquema de resfriamento linear:
 - $T = T_0 + \beta t$,
 - ou $T = T_0 + \delta t$,
 - onde T_0 é a temperatura inicial
 - e T é o tempo das iterações
 - β é a taxa de resfriamento deveria ser escolhido quando $T \rightarrow 0$ quando $t \rightarrow tf$ (ou o máximo número N de iterações) logo o
 - $\beta = (T_0 - T_f)/tf$,
 - De outra maneira o resfriamento geométrico essencialmente decrementa a temperatura por um fator $0 < \alpha < 1$ então o T é substituído por αT ou $T = T_0 \alpha^t$, $t=1,2,\dots,tf$

Simulated Annealing



- Parâmetros:
 - A vantagem do segundo método é que $T \rightarrow 0$ quando $t \rightarrow \infty$ por esta razão não precisa especificar o número de iterações.
 - O processo de resfriamento deveria ser lento o suficiente para permitir que o sistema estabilize facilmente. Na prática $\alpha=0.7 \sim 0.99$.
 - Para um dada temperatura, múltiplas ativações da função objetivo são necessárias. Se poucas avaliações são feitas, existe o perigo do sistema não estabilizar e consequentemente não convergir para sua otimalidade global.
 - Se muitas avaliações são feitas, consome muito tempo e o sistema convergirá muito devagar, pois o número de iterações para atingir estabilidade deve ser exponencial ao tamanho do problema.
 - Existe um equilíbrio fino entre o número de avaliações e a qualidade da solução.

Simulated Annealing



- Parâmetros:
 - Pode-se também fazer muitas avaliações em alguns níveis de temperatura ou fazer algumas avaliações em muitos níveis de temperatura.
 - Existem duas maneiras principais para setar o número de iterações: fixo ou variado.
 - O primeiro usa um número fixo de iterações em cada temperatura, enquanto o segundo pretende aumentar o número de iterações em temperaturas mais baixas e então o mínimo local pode ser completamente explorado.

Simulated Annealing



Simulação termodinâmica	Otimização
Estados do sistema	Soluções viáveis
Energia	Custo
Mudança de estado	Estrutura de vizinhança
Temperatura	Parâmetro de controle
Estado congelado	Solução heurística

Tabela 1: Analogia entre o sistema físico e o problema de otimização.

Fonte: Dowsland (1995, p. 23).

Simulated Annealing

- O Algoritmo Simulated Annealing:

Simulated Annealing Algorithm

Objective function $f(\mathbf{x})$, $\mathbf{x} = (x_1, \dots, x_p)^T$

Initialize initial temperature T_0 and initial guess $\mathbf{x}^{(0)}$

Set final temperature T_f and max number of iterations N

Define cooling schedule $T \mapsto \alpha T$, ($0 < \alpha < 1$)

while ($T > T_f$ and $n < N$)

Move randomly to new locations: $\mathbf{x}_{n+1} = \mathbf{x}_n + \epsilon$ (random walk)

Calculate $\Delta f = f_{n+1}(\mathbf{x}_{n+1}) - f_n(\mathbf{x}_n)$

Accept the new solution if better

if not improved

Generate a random number r

Accept if $p = \exp[-\Delta f/T] > r$

end if

Update the best \mathbf{x}_ and f_**

$n = n + 1$

end while



Simulated Annealing

- O Algoritmo Simulated Annealing (de outra referência):

```
procedure simulated annealing
begin
   $t \leftarrow 0$ 
  initialize  $T$ 
  select a current point  $v_c$  at random
  evaluate  $v_c$ 
  repeat
    repeat
      select a new point  $v_n$ 
        in the neighborhood of  $v_c$ 
      if  $eval(v_c) < eval(v_n)$ 
        then  $v_c \leftarrow v_n$ 
      else if  $random[0, 1) < e^{\frac{eval(v_n) - eval(v_c)}{T}}$ 
        then  $v_c \leftarrow v_n$ 
    until (termination-condition)
     $T \leftarrow g(T, t)$ 
     $t \leftarrow t + 1$ 
  until (halting-criterion)
end
```

Fig. 5.3. Structure of simulated annealing



Simulated Annealing

- Para encontrar o valor de T_0 , pode usar alguma informação sobre a função objetivo.
- Se já se sabe a máxima mudança da função objetivo $\max(\Delta f)$ pode ser usado para estimar o valor inicial da temperatura T_0 para uma dado valor de probabilidade p_0 :
- $$T_0 \approx \frac{-\max(\Delta f)}{\ln p_0}$$
- Se não se sabe a máxima mudança possível da função objetivo, uma abordagem heurística pode ser usada.



Simulated Annealing

- Pode-se iniciar as avaliações a partir de uma temperatura muito alta (onde quase todas as mudanças são aceitas) e reduzir a temperatura rapidamente até 50 a 60% dos piores movimentos aceitos.
- Então usar essa temperatura como a nova temperatura inicial.
- Para a temperatura final ela deveria ser zero na teoria e então nenhum movimento ruim pode ser aceito.
- Entretanto, se o $T_f \rightarrow 0$, mais avaliações não necessárias serão realizadas. Na prática $T_f = 10^{-10} \sim 10^{-15}$ dependendo da qualidade requerida da solução e das restrições de tempo.