Utrecht, 14 juni 2012

Modellen en Simulatie
Simulated Annealing

Gerard Sleijpen

Universiteit Utrecht
Department of Mathematics

http://www.staff.science.uu.nl/~sleij101/

Program

- Optimaliseren
- Het type probleem
- Het handelsreizigersprobleem.
- NP-compleet
- Simulated Annealing
- De kans op succes
- Hoe T te kiezen?
- Fysische interpretatie

Handelsreizigersprobleem

Gegevens. n steden $S_1, S_2, \ldots, S_n$. De afstand L_{ij} (over de weg) tussen ieder tweetal steden S_i en S_j is bekend.

Traveling salesman problem. Vind een rondreis met de kleinste totaal afstand die langs **alle** steden gaat.

Als $\pi : \{2, \ldots, n\} \to \{2, \ldots, n\}$ een permutatie is en

$$S_1 \to S_{\pi(2)} \to S_{\pi(3)} \to \ldots \to S_{\pi(n)} \to S_1$$

beschrijft de rondreis, dan is de totaal afstand voor deze reis:

$$\mathbf{f}(\pi) \equiv L_{1,\pi(2)} + L_{\pi(2),\pi(3)} + \ldots + L_{\pi(n-1),\pi(n)} + L_{\pi(n),1}.$$

Karakteristieken: alle afstanden L_{ij} zijn bekend. Gegeven een rondreis (permutatie π) dan is $\mathbf{f}(\pi)$ **zeer** snel uit te rekenen.

Brute-force

Met $n = 31$ steden is het aantal permutaties

$$30! \approx 2.6\,10^{30}$$

Stel dat je 10^9 (1 Terra) mogelijke rondreizen per seconde kunt checken, dan duurt de brute-force methode

$$\approx 2.6\,10^{21}\,\text{sec} \approx 8\,10^{13}\,\text{jaar}.$$

(1 jaar $\approx 3\,10^7$ sec.).

[Wikipedia] In the theory of computational complexity, the decision version of TSP belongs to the class of NP-complete problems. Thus, it is assumed that there is no efficient algorithm for solving TSP problems. In other words, it is likely that the worst case running time for any algorithm for TSP increases exponentially with the number of cities, so even some instances with only hundreds of cities will take many CPU years to solve exactly.

NP-compleet

Een probleem is van complexiteitsklasse **Nondeterministic Polynomial time** als "gemakkelijk" na te gaan is dat een oplossing correct is.

Voorbeeld [het deelverzameling probleem]. Zij $\mathcal{F}$ een eindige collecte gehele getallen (vb $\mathcal{F} = \{-7, -3, -2, 5, 8\}$). Bepaal, indien mogelijk, een deelverzameling $\mathcal{G}$ van $\mathcal{F}$ waarvan alle getallen optellen tot 0 (als $\mathcal{G} = \{-3, -2, 5\}$)?

"Gemakkelijk" Benodigde rekentijd (of aantal rekenoperaties) is $\sim n^k$, waarbij n het aantal input parameters is en k een positief geheel getal. (**polynomiale tijd**)

Een NP-probleem is **NP-compleet** als er geen proces <u>**bekend**</u> is dat een oplossing vindt in polynomiale tijd.

Handelsreizigersprobleem

Brute-force werkt niet.

Er zijn diverse 'intelligentere' exacte oplosmethoden bekend die de oplossing vinden in $\sim n^2\, 2^n$ en zelfs in $\sim 1.9999^n$

Alternatief. Tevreden zijn als π_0 'vlug' te berekenen is met π_0 bijvoorbeeld zo dat

$$\mathbf{f}(\pi_0) \leq 1.2 \min_\pi \mathbf{f}(\pi)$$

Alternatief. Tevreden zijn als π_0 'vlug' te berekenen is met π_0 zo dat met grote waarschijnlijkheid

$$\mathbf{f}(\pi_0) \approx \min_\pi \mathbf{f}(\pi)$$

Simulated annealing

Simulated annealing

$\mathbf{f} : \mathcal{V} \to \mathbb{R}$. Vind $\mathbf{x}_0 \in \mathcal{V}$ zodat $\mathbf{f}(\mathbf{x}_0) \leq \mathbf{f}(\mathbf{x})$ alle $\mathbf{x} \in \mathcal{V}$.

Vb. 1) $\mathcal{V} = \{\pi \text{ permutatie van } \mathbf{N} \equiv \{1, 2, \ldots, n\}\}$,

$$\mathbf{f}(\pi) = L_{1,\pi(2)} + \textstyle\sum_{j=2}^{n-1} L_{\pi(j),\pi(j+1)} + L_{\pi(n),1}$$

2) $\mathcal{V} = \{t_k = \frac{k-1}{N-1} \mid k = 1, \ldots, N\}$, $\quad \mathbf{f}(t_k) = f_k$.

Voor $\mathbf{x} \in \mathcal{V}$ zij $\mathcal{B}(\mathbf{x}) \subset \mathcal{V}$ een set van '**buur**' punten van $\mathbf{x}$.

Als $\mathbf{x}$ een buur is van $\mathbf{y}$ dan is $\mathbf{y}$ een buur van $\mathbf{x}$.

vb 1) π_1 is een buur van π als
er $j_1, j_2 \in \mathbf{N}$ zijn $j_1 \neq j_2$ zo dat
$\pi_1(j_1) = \pi(j_2)$, $\pi_1(j_2) = \pi(j_1)$
en $\pi_1(i) = \pi(i)$ voor de andere $i \in \mathbf{N}$.

vb 2) t_k is een buur van t_m als, voor $M = 10$,
$|n - j| \leq M$ voor 'n k met $m = j \bmod N$

Simulated annealing

$\mathbf{f} : \mathcal{V} \to \mathbb{R}$. Vind $\mathbf{x}_0 \in \mathcal{V}$ zodat $\mathbf{f}(\mathbf{x}_0) \leq \mathbf{f}(\mathbf{x})$ alle $\mathbf{x} \in \mathcal{V}$.

Vb. 1) $\mathcal{V} = \{\pi \text{ permutatie van } \mathbf{N} \equiv \{1, 2, \ldots, n\}\}$,

$$\mathbf{f}(\pi) = L_{1,\pi(2)} + \textstyle\sum_{j=2}^{n-1} L_{\pi(j),\pi(j+1)} + L_{\pi(n),1}$$

2) $\mathcal{V} = \{t_k = \frac{k-1}{N-1} \mid k = 1, \ldots, N\}$, $\quad \mathbf{f}(t_k) = f_k$.

Voor $\mathbf{x} \in \mathcal{V}$ zij $\mathcal{B}(\mathbf{x}) \subset \mathcal{V}$ een set van '**buur**' punten van $\mathbf{x}$.

Kies $\mathbf{x}$ en $T > 0$
Herhaal
1) Kies random $\mathbf{y} \in \mathcal{B}(\mathbf{x})$
2) Als $\mathbf{f}(\mathbf{y}) < \mathbf{f}(\mathbf{x})$ dan $\mathbf{x} \leftarrow \mathbf{y}$
anders $\mathbf{x} \leftarrow \mathbf{y}$ met kans $e^{(\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{y}))/T}$

Simulated annealing

$\mathbf{f} : \mathcal{V} \to \mathbb{R}$. Vind $\mathbf{x}_0 \in \mathcal{V}$ zodat $\mathbf{f}(\mathbf{x}_0) \le \mathbf{f}(\mathbf{x})$ alle $\mathbf{x} \in \mathcal{V}$.

Vb. 1) $\mathcal{V} = \{\pi \text{ permutatie van } \mathbf{N} \equiv \{1, 2, \ldots, n\}\}$,

$$\mathbf{f}(\pi) = L_{1,\pi(2)} + \sum_{j=2}^{n-1} L_{\pi(j),\pi(j+1)} + L_{\pi(n),1}$$

2) $\mathcal{V} = \{t_k = \frac{k-1}{N-1} \mid k = 1, \ldots, N\}$, $\quad \mathbf{f}(t_k) = f_k$.

Voor $\mathbf{x} \in \mathcal{V}$ zij $\mathcal{B}(\mathbf{x}) \subset \mathcal{V}$ een set van '**buur**' punten van $\mathbf{x}$.

> Kies $\mathbf{x}$ en $T > 0$
> Herhaal
> 1) Kies random $\mathbf{y} \in \mathcal{B}(\mathbf{x})$
> 2) Als $\mathbf{f}(\mathbf{y}) < \mathbf{f}(\mathbf{x})$ dan $\mathbf{x} \leftarrow \mathbf{y}$
> anders $\mathbf{x} \leftarrow \mathbf{y}$ met kans $e^{(\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{y}))/T}$

Wat is de kans dat we $\quad \mathbf{x}_0 \equiv \mathrm{argmin}_{\mathbf{x}}\, \mathbf{f}(\mathbf{x}) \quad$ *vinden?*

> Kies $\mathbf{x}$ en $T > 0$
> Herhaal
> 1) Kies random $\mathbf{y} \in \mathcal{B}(\mathbf{x})$
> 2) Als $\mathbf{f}(\mathbf{y}) < \mathbf{f}(\mathbf{x})$ dan $\mathbf{x} \leftarrow \mathbf{y}$
> anders $\mathbf{x} \leftarrow \mathbf{y}$ met kans $e^{(\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{y}))/T}$

Model. Zij $p_{\mathbf{x},\mathbf{y}}$ de kans dat $\mathbf{x}$ vervangen wordt door $\mathbf{y}$.

Zij b het aantal buren in $\mathcal{B}(\mathbf{x})$

(neem aan dit aantal is hetzelfde voor iedere $\mathbf{x}$.)

Dan, $\quad$ als $\mathbf{y} \in \mathcal{B}(\mathbf{x})$

$$p_{\mathbf{x},\mathbf{y}} = \begin{cases} \frac{1}{b} & \text{als} \quad \mathbf{f}(\mathbf{y}) < \mathbf{f}(\mathbf{x}) \\ \frac{1}{b} e^{(\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{y}))/T} & \text{als} \quad \mathbf{f}(\mathbf{y}) \ge \mathbf{f}(\mathbf{x}) \end{cases}$$

als $\mathbf{y} \notin \mathcal{B}(\mathbf{x})$, $\mathbf{y} \ne \mathbf{x}$,

$$p_{\mathbf{x},\mathbf{y}} = 0$$

$p_{\mathbf{x},\mathbf{x}} = 1 - \sum p_{\mathbf{x},\mathbf{y}}$ waarbij gesommeerd over alle $\mathbf{y} \ne \mathbf{x}$.

vb 2) $\mathcal{V} = \{t_k = \frac{k-1}{N-1} \mid k = 1, \ldots, N\}$ met $N = 6$
en $\mathcal{B}(t_k) = \{t_{k-1}, t_{k+1}\}$ (de index mod N).

$\mathbf{p}_n(k)$ is de kans dat we in de n-de slag t_k vinden.

$$\mathbf{P} = \begin{bmatrix} * & \frac{1}{2}e^{(f_2 - f_1)/T} & 0 & 0 & 0 & \frac{1}{2} \\ \frac{1}{2} & * & \frac{1}{2} & 0 & 0 & 0 \\ 0 & \frac{1}{2}e^{(f_2 - f_3)/T} & 0 & \frac{1}{2}e^{(f_4 - f_3)/T} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & * & \frac{1}{2} & 0 \\ 0 & 0 & 0 & \frac{1}{2}e^{(f_4 - f_5)/T} & * & \frac{1}{2} \\ \frac{1}{2}e^{(f_1 - f_6)/T} & 0 & 0 & 0 & \frac{1}{2}e^{(f_5 - f_6)/T} & 0 \end{bmatrix}$$

$$\mathbf{p}_n = \mathbf{P}^n \mathbf{p}_0.$$

$\mathbf{P}$ is irreducibel en a-periodiek. Dus: 1 is de dominante eigenwaarde en is $\mathbf{q}$ de kans eigenvector bij 1, dan $\mathbf{p}_n \to \mathbf{q}$. Op den duur is (ongeacht de start) de kans dat we t_k vinden gelijk aan $\mathbf{q}(k)$.

De dominante kans eigenvector

Stelling. Zij $\mathbf{P}$ een $N \times N$ irreducibele, a-periodieke kansmatrix. Zij $\mathbf{q}$ een kansvector van dimensie N. Als

$$(*) \qquad p_{i,j}\, q_j = p_{j,i}\, q_i \qquad (i, j = 1, \ldots, N),$$

dan is $\mathbf{q}$ de dominante kans eigenvector van $\mathbf{P}$.

Opmerking. Met $\mathbf{D} \equiv \mathrm{diag}(q_1, \ldots, q_N)$ geldt dat

$$(*) \quad \Leftrightarrow \quad \mathbf{PD} = \mathbf{DP}^\top:$$

$\mathbf{PD}$ is symmetrisch $\Leftrightarrow (*)$.

Bewijs. We bewijzen dat $\mathbf{q}$ een eigenvector is bij eigenwaarde 1 als $(*)$ geldt. We gebruiken dat $\mathbf{1}^\top\mathbf{P} = \mathbf{1}^\top$ voor een kans matrix $\mathbf{P}$ en dat voor een irreducible, a-periodiek kansmatrix de dominante kans eigenvector uniek is.

$$\mathbf{Pq} = \mathbf{PD1} = \mathbf{DP}^\top\mathbf{1} = \mathbf{D}(\mathbf{1}^\top\mathbf{P})^\top = \mathbf{D1} = \mathbf{q}.$$

De dominante kans eigenvector

Stelling. $p_{i,j}\, q_j = p_{j,i}\, q_i \;\forall i, j \quad \Rightarrow \quad \mathbf{P}^n \mathbf{p}_0 \to \mathbf{q}\ (n \to \infty)$.

Stelling. Met $q(\mathbf{x}) = \frac{1}{\sigma} e^{-\mathbf{f}(\mathbf{x})/T}$ en $\sigma \equiv \sum_{\mathbf{x} \in \mathcal{V}} e^{-\mathbf{f}(\mathbf{x})/T}$

is $q(\mathbf{x})$ de kans dat het proces op den duur $\mathbf{x}$ oplevert.

Opmerking. Om tot deze conclusie te komen hebben we de matrix $\mathbf{P}$ nodig (en de $p_{\mathbf{x},\mathbf{y}}$), maar $\mathbf{P}$ hebben we **niet** expliciet nodig om het proces uit te voeren.

Als het minimum in precies in een $\mathbf{x}$ wordt aangenomen, zeg in $\mathbf{x} = \mathbf{x}_0$, dan $q(\mathbf{x}_0) \to 1$ als $T \to 0$.

Stelling. Voor kleine T is de kans zeer groot dat het proces op den duur $\mathbf{x}_0$ vindt.

Hoe kiezen we T?

+ T klein dan is de kans groot dat we op den duur in een minimaliserende $\mathbf{x}$ terecht komen.

– T klein dan duurt ‘op den duur’ heel lang

Praktische strategie.
Begin met grotere T, verklein T na een aantal slagen.

Met grotere T hebben we al ‘vlug’ een grotere kans dat we in de kleinere $\mathbf{f}$-waarden terecht komen (maar er is ook nog een redelijke kans dat het proces andere punten oplevert). Die vormen een goed uitgangspunt voor de kleinere T, waarbij de kans dat het proces ‘echt’ verkeerde punten oplevert kleiner is.

Met welke strategie moeten we T verkleinen?
Getemperd (annealed) Trial and error.