

Martin Janžura, ÚTIA ČSAV, Praha

Pravděpodobnostní algoritmus nazývaný pro určitou fyzikální analogii "simulovaným žiháním" je vhodný pro řešení optimalizačních úloh na rozsáhlých diskrétních množinách. Tento algoritmus bude stručně popsán a demonstrován na úloze klasifikace obrazu.

1. Úvod

Nalézt maximum obecné funkce na diskrétní množině lze v zásadě jediné prohledáním. Je-li tato diskrétní množina příliš rozsáhlá, nelze ji prohledávat bod po bodu, ale je nutné postupovat namátkově. Takový postup pak nazveme pravděpodobnostním algoritmem.

Hledat však namátkově v celé veliké množině by také nevedlo rychle k cíli. Je účelnější omezit se vždy na nějaké okolí předchozího bodu, zvláště jestliže lze funkční hodnoty v takovém okolí snadno porovnávat. Toto omezení hrozí možností, že nás "lokální" změny dovedou do pouze lokálního maxima. Tomuto nebezpečí metoda čelí svou podstatou, která připouští náhodně i změny "k horšímu", ale umožňující opustit slepou uličku vedoucí k lokálnímu optimu.

2. Algoritmus

Mějme tedy konečnou množinu Ω a na ní definovanou reálnou funkci $F: \Omega \rightarrow \mathcal{R}$. Cílem je nalézt maximum funkce F .

Jestliže po k iteracích jsme v bodě ω_k^* , pak v dalším kroku vybereme rovnoměrně náhodně

$$\tilde{\omega} \in \partial_{i(k)}(\omega_k^*),$$

kde $\partial_{i(k)}(\omega_k^*)$ je nějaké okolí bodu ω_k^* , přičemž řekněme, že máme N možností $\partial_1, \dots, \partial_N$, jak takové okolí konstruovat, a tyto možnosti nějakým způsobem (třeba periodicky) střídáme, tedy $i(k) \in \{1, \dots, N\}$.

Označíme-li $\Delta_k = F(\tilde{\omega}) - F(\omega_k^*)$, pak dosadíme

$$\omega_{k+1}^* := \tilde{\omega}$$

s pravděpodobností $\exp\left\{\min\left\{0, \frac{1}{T(k)}\Delta_k\right\}\right\}$. S doplňkovou pravděpodobností pak ponecháme

$$\omega_{k+1}^* := \omega_k^* .$$

Zde $T(k) \rightarrow 0$ pro $k \rightarrow \infty$.

Označme také $M = \arg \max F \subset \Omega$ množinu těch bodů, v nichž se nabude maxima funkce F .

Platí zhruba takovéto tvrzení:

Jestliže

$T(k) \approx 1/\log(k)$ a $\{1, \dots, N\} \subset \{i(k), \dots, i(k+\tau)\}$ pro nějaké pevné τ a každé k , pak

$$P(\omega_k^* \in M) \rightarrow 1 \quad \text{pro } k \rightarrow \infty .$$

Naznačíme stručně, jak tvrzení dokazovat.

Každý krok iterace je řízen vlastně nějakou maticí pravděpodobností přechodu

$$P_{i(k)}^{T(k)}(\tilde{\omega}|\omega),$$

která závisí na výběru typu okolí $i(k) \in \{1, \dots, N\}$ a reálném parametru $T(k)$.

Nejprve se ukáže, že při fixním T a "dostatečně pravidelném" výběru $i(k)$ platí

$$\prod_{k=1}^K P_{i(k)}^T \rightarrow P^T \quad \text{pro } K \rightarrow \infty,$$

kde

$$P^T(\omega) \approx \exp\left\{\frac{1}{T} F(\omega)\right\}.$$

Snadno je vidět, že $P^T \rightarrow P^0$ pro $T \rightarrow 0$, kde

$$P^0(\omega) = \frac{1}{|M|} \delta_M(\omega)$$

je rovnoměrné rozdělení na bodech maxima.

Je tedy třeba nalézt takové tempo klesání $T(k) \rightarrow 0$, aby přímo

$$\prod_{k=1}^K P_{i(k)}^{T(k)} \rightarrow P^0.$$

Ukazuje se, že postačující tempo je nepřímo úměrné logaritmu počtu iterací, tedy

$$T(k) = c / \log(k).$$

Celý důkaz je samozřejmě technicky náročný a lze jej ve speciálním případě nalézt např. v [1].

3. Typická situace

Metoda se uplatní zejména v situacích, kdy $\Omega = \prod_{v \in V} \Omega_v$ je mnohorozměrný kartézský součin.

Potom obvykle volíme lokální okolí ve tvaru

$$\partial_v(\omega^*) = \left\{ \omega \in \Omega : \omega_t = \omega_t^*, t \neq v \right\}$$

pro každé $v \in V$.

Jestliže navíc kriteriální funkce F je dána jako suma

$$F(\omega) = \sum_{A \in \mathcal{A}} F_A(\omega_A)$$

kde F_A závisí pouze na souřadnicích z množiny A , potom lze s výhodou vyčíslivat i lokální změny

$$\Delta_v = F(\omega) - F(\omega^*) = \sum_{A \ni v} \left(F_A(\omega_A) - F_A(\omega_A^*) \right)$$

v každém kroku iterace.

4. Příklady

A. Postup lze využít v klasických optimalizačních úlohách typu "problému obchodního cestujícího", zejména v algoritmech založených na lokálních změnách (viz [2]).

B. Veliké možnosti leží v oblasti "expertních systémů", pokud se zde poněkud změní filozofie. Doposud se úloha stavěla tak, že je třeba nalézt pravděpodobnosti (váhy) všech jednotlivých jevů nebo výroků. Jestliže však připustíme, že nás zajímají především jevy a výroky s největší pravděpodobností (vahou) za daných podmínek, uvedená metoda se přímo nabízí.

Rekneme, že každá konfigurace $\omega_V \in \Omega$ má určitou pravděpodobnost $P(\omega_V)$, která se řídí logaritmicko lineárním modelem, tj.

$$L(\omega_V) = \log P(\omega_V) = \sum_{A \in \mathcal{A}} U_A(\omega_A) + C$$

kde $\mathcal{A} \subset \exp V$ je nějaký systém podmnožin množiny V a

$U_A(\omega_A)$ je interakce příslušná konfiguraci ω_A příznaků z množiny $A \in \mathcal{A}$.

Ti, kteří odmítají pravděpodobnostní přístup, mohou interpretovat hodnoty $U_A(\omega_A)$ jako příslušné váhy a $L(\omega_V)$ jako jakousi globální váhu konfigurace ω_V .

Konkrétní hodnoty $U_A(\omega_A)$ jsou považovány za známé, získané buď od expertů nebo ze statistického materiálu.

Úloha pak zní takto: Jestliže známe konfiguraci ω_V^* (napozorované příznaky), kde $V_1 \subset V$, chci nalézt takovou konfiguraci $\omega_{V \setminus V_1}^*$, pro kterou platí

$$L(\omega_V^*) = \max_{\omega_{V \setminus V_1}} L(\omega_{V \setminus V_1}, \omega_{V_1}^*)$$

Použijí tedy uvedený algoritmus. Výhoda metody je nesporná: Nemám žádné problémy s konzistencí expertních znalostí a nemám žádné problémy s jejich agregací. Vyhnou se tedy bez ztráty obecnosti modelu těm problémům, se kterými standardní přístupy těžce zápolí.

C. Budeme nyní metodu demonstrovat na jednoduché úloze z oblasti "zpracování obrazu". Necht' $\omega_V \in \{0,1\}^V = \Omega$, $V = 40 \times 40 \subset \mathbb{Z}^2$ je "dvojbarevný obraz", který však pozorujeme aditivně zašumělý bílým gaussovským šumem, tj. pozorujeme

$$y_V = \omega_V + x_V,$$

kde x_t , $t \in V$ jsou nezávislé s rozdělením $N(0, \sigma^2)$.

Otázka je tedy, jaký byl původní obraz ω_V , jestliže jsme napozorovali y_V .

Budeme maximalizovat aposteriorní pravděpodobnost, tedy

$$\max_{\omega_V} L(\omega_V | y_V),$$

kde

$$L(\omega_V | y_V) = \log P(\omega_V | y_V) = \sum_{t \in V} \log P(y_t | \omega_t) + \log P(\omega_V) - \log P(y_V).$$

Přitom $\log P(y_t | \omega_t) = -\frac{(y_t - \omega_t)^2}{2\sigma^2} - \log(\sqrt{2\pi}\sigma)$ pro každé $t \in V$

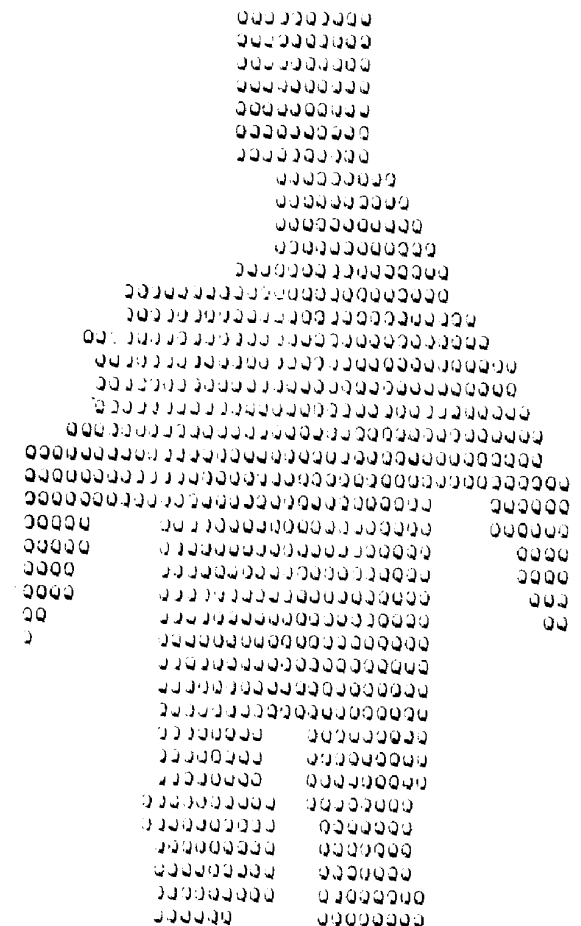
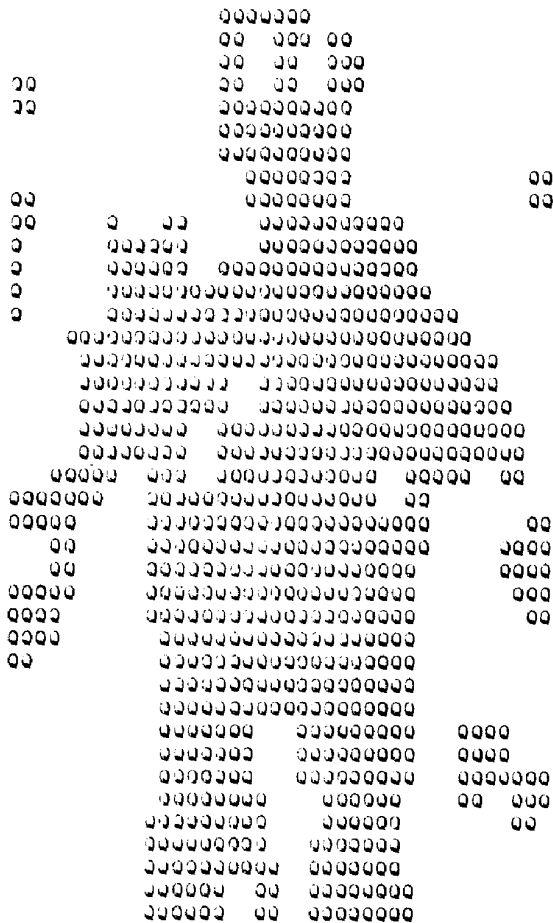
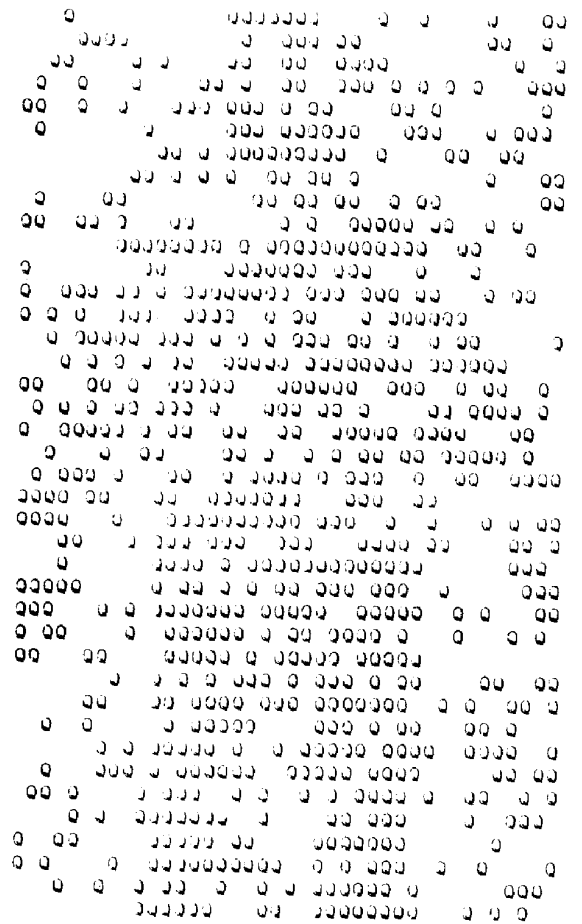
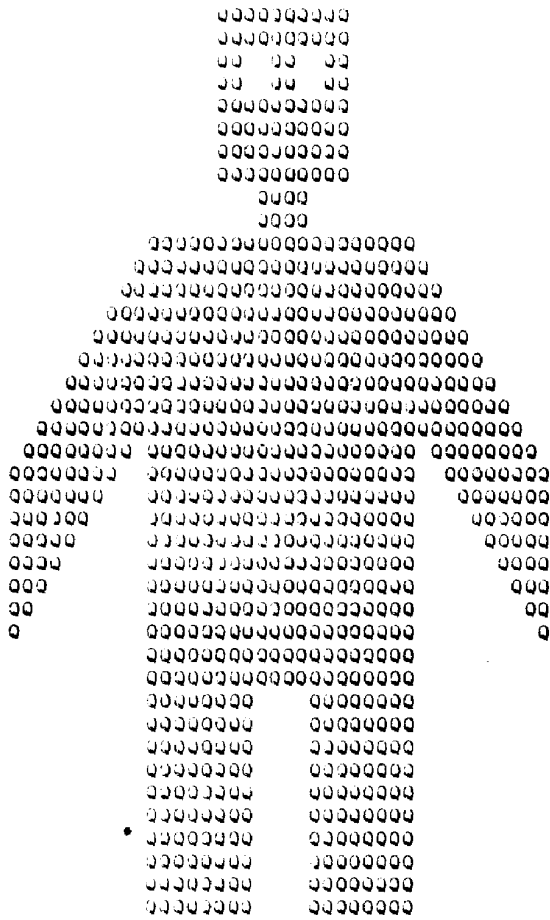
a $P(\omega_V)$ je apriorní pravděpodobnost, o které budeme předpokládat, že je markovská (Isingův model), tedy

$$\log P(\omega_V) = \lambda_0 \cdot \sum_{v \in V} \omega_v + \lambda_1 \cdot \sum_{\|v-t\|=1} \omega_v \omega_t + C.$$

Dále budeme předpokládat, že $\lambda_1 > 0$ a $\lambda_0 = -2\lambda_1$, což znamená, že stejné hodnoty mají snahu spolu sousedit (vytvořit homogenní plochy) a přitom jsou obě barvy stejně pravděpodobné.

Jestliže dosadíme, vynecháme aditivní konstanty a upravíme, obdržíme úlohu maximalizovat funkci

$$\tilde{L}(\omega_V | y_V) = \sum_V \omega_v (y_v - \frac{1}{2}) + \sigma^2 \lambda_1 \left[\sum_{\|t-v\|=1} \omega_t \omega_v - 2 \sum_V \omega_v \right].$$



Postup jsme odzkoušeli na obrázku 1, který je sice deterministický, ale budeme ho považovat za generovaný Isingovým modelem. Tento obrázek jsme zašuměli bílým gaussovským šumem s rozptylem $\sigma^2 = 1$. Na dalších obrázcích jsou pak výsledky optimalizace, které závisejí na volbě (neznámého) parametru $B = \sigma^2 \cdot \lambda_1$ a "teplotě" $T(k)$. Volba $T(k) = 0$ znamená striktní rozhodování "na tvrdo", bez náhodného žihání. Při $T(k) = 0$ novou hodnotu přijmu jen tehdy, když přináší zlepšení.

Výsledky ukazují, že rozhodující pro správné řešení je sice vhodná volba parametru L (volba $L = 0$ na obr. 2 předpokládá nezávislost, která jistě v původním obrázku není), ale porovnáním obr. 3 a 4 vidíme, že "náhodné žihání" odstranilo z řešení falešné shluky a vedlo k lepším výsledkům.

Toto je skutečně velmi jednoduchý příklad. Obdobně by bylo možné řešit i mnohem složitější úlohy s více barvami, složitější apriorní distribucí, závislým šumem apod.

4. Závěr

Metoda "simulovaného žihání" je skutečně velmi účinná a s rozvojem výpočetní techniky (zejména paralelismu) lze očekávat její nejširší použití.

Literatura

- [1] Gemam, S., Gemam, D.: Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE, Vol. PAMI-6, No. 6(1984), 721-741.
- [2] Černý, V.: A thermodynamical approach to the travelling salesman problem: an efficient simulation algorithm. Preprint, Inst. Phys. & Biophys., Comenius Univ. Bratislava, 1982.