4. prednáška

Teória genetického algoritmu

Teória GA bola vypracovaná J.Hollandom (1975)**.**

Schéma je reťazec dĺžky *k* zostrojený nad množinou symbolov {0,1,#}, formálne schéma $\sigma \in \{0,1,\#\}^k$, kde "#" je tzv. divoký symbol (wildcard).

Hovoríme, že reťazec $\alpha \in \{0,1\}^k$ je **príkladom** schémy $\sigma \in \{0,1,\#\}^k$, formálne $\alpha \in \sigma$, vtedy, ak v každej polohe schémy σ s iným ako divokým symbolom je hodnota znaku reťazca α totožná so znakom v σ.

$$\alpha \in \sigma \Leftrightarrow \left(\forall i \in \{1,2,...,k\} : \sigma_i \in \{0,1\} \Rightarrow \sigma_i = \alpha_i\right)$$

Binárny reťazec dĺžky k=7 α=(1101011) je príkladom schémy σ=(11##0##)

$$\left.\begin{array}{l}\sigma=(11\#\#0\#\#)\\ \alpha=(1101011)\end{array}\right\} \longrightarrow \alpha\in\sigma$$

Množina $I(\sigma)$ obsahuje všetky možné príklady α schémy σ

$$I(\sigma)=\{\alpha;\alpha\in\sigma\}$$

Pre schéma σ=(11##0##) množina $I(\sigma)$ obsahuje 16 binárnych reťazcov, ktoré môžu byť zostrojené tak, že divoké symboly "#" systematicky nahradíme "0" alebo "1"

$$I(\sigma)=\{(1100000),(1100001),\ldots,(1111011)\}$$

Rád schémy σ, označený o**(σ),** popisuje počet symbolov "0" alebo "1" v σ

$$o(\sigma) = \left|\{i; \sigma_i \in \{0,1\}\}\right|$$

Pre σ=(11∗∗0∗∗) platí o(σ)=3, pretože schéma obsahuje len tri "nedivoké" symboly.

Dĺžka Δ schémy σ je určená ako rozdiel medzi indexom posledného binárneho symbolu v σ a prvým binárnym symbolom v σ

$$\Delta(\sigma) = \max\{i; \sigma_i \in \{0,1\}\} - \min\{i; \sigma_i \in \{0,1\}\}$$

Pre σ=(11∗∗0∗∗) platí Δ(σ)=5-1=4.

Pravdepodobnosti prežitia schémy **σ pri aplikácii mutácie**

Keď mutácia je určená s pravdepodobnosťou jedno bitovej mutácie P_{mut}, potom pravdepodobnosť prežitia schémy je

$$(1-P_{mut})^{o(\sigma)}$$

(1) Pravdepodobnosť prežitia jedného bitu je ($1-P_{\mathrm{mut}}$).
(2) Schéma σ obsahuje práve o(σ) binárnych symbolov.

V mnohých praktických aplikáciách genetického algoritmu pravdepodobnosť P_{mut} je veľmi malá, preto pravdepodobnosť prežitia schémy môžeme zjednodušiť

$$1-o(\sigma)P_{mut}$$

Pravdepodobnosti prežitia schémy **σ pri aplikácii kríženia**

- Schéma σ zaniká aplikovaním kríženia s pravdepodobnosťou $P_{cross}\Delta(\sigma)/(k-1)$.
- Každé kríženie s chromozómom $\alpha\in P$, ktorý je príkladom schémy σ, $\alpha\in\sigma$, automaticky garantuje zachovanie tejto schémy. Nech $N(\sigma)=\left|I(\sigma)\cap P\right|$ popisuje počet príkladov schémy σ v populácii *P*, potom len frakcia $1-N(\sigma)/\left|P\right|$ z populácie *P* môže spôsobovať deštrukciu schémy σ krížením.

Minimálna pravdepodobnosť prežitia schémy σ v populácii *P* je určená vzťahom

$$1-P_{cross}\frac{\Delta(\sigma)}{k-1}\left(1-\frac{N(\sigma)}{\left|P\right|}\right)\approx 1-P_{cross}\frac{\Delta(\sigma)}{k-1}$$

(zanedbali sme súčinový člen)

Problém výberu chromozómov vstupujúcich do reprodukcie

Strednú silu schémy σ je určená vzťahom

$$F(\sigma) = \frac{1}{N(\sigma)} \sum_{\alpha \in I(\sigma) \cap P} F(\alpha)$$

Stredná sila chromozómov z populácie P

$$\overline{F} = \frac{1}{|P|} \sum_{\alpha \in P} F(\alpha)$$

Pri prechode z generácie t na nasledujúcu generáciu t+1 v GA, ktorý obsahuje len výber chromozómov podľa rulety a reprodukčný proces je plne ignorovaný, nastane nasledujúca zmena počtu výskytu schémy σ v populácii

$$N_{t+1}(\sigma) = N_t(\sigma)\frac{F_t(\sigma)}{\overline{F}_t} \qquad (\clubsuit)$$

Nech σ je schéma, ktorej stredná hodnota sily vzhľadom k strednej hodnote chromozómov je určená vzťahom

$$F_t(\sigma) = \overline{F}_t \times (1+c)$$

kde c je malé kladné alebo záporné číslo. Opakovaným použitím tohto vzťahu pre počiatočný stav t=0 dostaneme "idealizovanú" rovnicu

$$F_t(\sigma) = F_0(\sigma) \times (1+c)^t$$

Záver: Pre proporcionálny výber (realizovaný ruletou) stredná hodnota síl schémy *σ exponenciálne rastie* (klesá) v priebehu mnohých generácií genetického algoritmu pre 1+c>1 (resp. 1+c<1).

Kombinovaním rovnice (♣), ktorá vyjadruje zmenu výskytu schémy σ v dôsledku proporcionálneho výberu, s pravdepodobnosťami prežitia schémy σ v dôsledku mutácie a kríženia, dostaneme finálnu formulu pre odhad počtu príkladov schémy pre generáciu t+1 populácie P za predpokladu, že reprodukčný proces obsahuje mutáciu, kríženie a výber pomocou rulety

$$N_{t+1}(\sigma) \approx N_t(\sigma)\frac{F_t(\sigma)}{\overline{F}_t}\left[1 - P_{cross}\frac{\Delta_t(\sigma)}{k-1}\times\left(1 - \frac{N_t(\sigma)}{|P|}\right)\right]\times (1 - P_{mut})^{o_t(\sigma)}$$

Jej zjednodušená verzia má tvar (zanedbajú sa "krížové" súčiny v hranatej zátvorke)

$$N_{t+1}(\sigma) \approx N_t(\sigma)\frac{F_t(\sigma)}{\overline{F}_t}\left[1 - P_{cross}\frac{\Delta_t(\sigma)}{k-1} - P_{mut}o_t(\sigma)\right]$$

Veta (Schema Theorem, Holland, 1975). Nech pre populáciu P_t platí, že $N_t(\sigma)$ je počet chromozómov ktoré obsahujú schému σ, $F_t(\sigma)$ je stredná hodnota sily chromozómov obsahujúcich schému σ a $\bar{F}_t$ je stredná hodnota sily všetkých chromozómov. Potom očakávaný počet chromozómov v nasledujúcej populácii, ktoré obsahujú schému σ je zdola ohraničený nerovnosťou

$$N_{t+1}(\sigma) \approx N_t(\sigma)\frac{F_t(\sigma)}{\bar{F}_t}(1-P_{zánik})$$

kde $P_{\text{zánik}}$ je pravdepodobnosť zániku schémy v dôsledku mutácie a kríženia je určená vzťahom,

$$P_{zánik} = P_{cross}\frac{\Delta_t(\sigma)}{k-1} - P_{mut}o_t(\sigma)$$

Podľa tejto vety, ak $F_t(\sigma) > \bar{F}_t$ (alebo $F_t(\sigma) < \bar{F}_t$), potom schéma σ sa vyskytuje v nasledujúcej generácii v chromozómoch s rastúcou (klesajúcou) frekvenciou. Tieto jednoduché závery sú založené na predpoklade, že pravdepodobnosť zániku schémy $P_{zánik}$ je malé číslo. Toto je obvykle splnené pre krátke schémy (s malým $\Delta(\sigma)$), nazývané **stavebné kamene** (building blocks).

Veta tvorí jeden z hlavných teoretických výsledkov genetického algoritmu, ktorý bol dosiahnutý Hollandom v r. 1975. Naznačuje mechanizmus práce tohto algoritmu menovite v prvých iteračných krokoch, kde schémy - stavebné kamene hrajú dôležitú úlohu.

Teoretická možnosť GA nájsť globálne minimum sú študované až v poslednej dobe. Uvedieme teorém, ktorý špecifikuje za akých podmienok je genetický algoritmus potenciálne schopný poskytnúť globálne minimum s jednotkovou pravdepodobnosťou.

Veta. Nech v genetickom algoritme postupnosť populácií P_0, P_1, ... je monotónna, t.j. pre každú generáciu t platí

$$\max_{\alpha \in P_{t+1}} F(\alpha) \geq \max_{\alpha \in P_t} F(\alpha)$$

potom genetický algoritmus asymptoticky pre $t \to \infty$ dosiahne globálneho optima

$$\alpha_{opt} = \lim_{t \to \infty} \left(\arg \min_{\alpha \in P_t} F(\alpha) \right)$$

Ako dosiahnuť *monotónnosť* genetického algoritmu? Toto môže byť jednoducho dosiahnuté pomocou tzv. *elitizmu*, ktorý je založený na jednoduchej myšlienke, že najlepšie riešenie z populácie P_t je automaticky prenesené aj do nasledujúcej populácie P_{t+1}.