

Wstęp do algorytmów ewolucyjnych

deterministic sampling - badanie statystycznej istotności metody reprodukcji - zarys projektu

Albert Sadowski, Krzysztof Wodecki
Wydział Matematyki i Nauk Informatycznych
Politechnika Warszawska

17 kwietnia 2013

1 Wstęp

Selekcja ukierunkowuje algorytm ewolucyjny w stronę lepszych rozwiązań. Z jednej strony, w fazie reprodukcji, wybierani są rodzice dla operacji genetycznych, z drugiej, w fazie sukcesji, tworzy się nową populację bazową na podstawie osobników znajdujących się w populacji potomnej i starej populacji bazowej. *Deterministic sampling* jest jedną z metod reprodukcji, którą porównamy z inną metodą reprodukcji - ruletkową reprodukcją proporcjonalną.

1.1 Reprodukacja ruletkowa

Zgodnie z [1], w reprodukcji proporcjonalnej definiuje się zmienną losową, określając dla każdego osobnika w populacji P^t prawdopodobieństwo jego reprodukcji według wzoru:

$$p_r(X) = \frac{\Phi(X)}{\sum_{Y \in P^t} \Phi(Y)} \quad (1)$$

gdzie $\Phi(X)$ - funkcja przystosowania. Prawdopodobieństwo wylosowania osobnika jest wprost proporcjonalne do jego wartości funkcji przystosowania.

1.2 Deterministic sampling

Deterministic sampling to także metoda reprodukcji proporcjonalnej ([2]). Wartość funkcji dopasowania każdego osobnika w populacji jest dzielona przez średnią wartość funkcji dopasowania w populacji:

$$v_d(X) = \lfloor \frac{\Phi(X)}{\sum_{Y \in P^t} \Phi(Y) / P^t} \rfloor \quad (2)$$

Jedynie część całkowita otrzymanej wartości jest brana pod uwagę. Jeżeli wartość v_d jest równa bądź większa od jedności osobnik jest wybierany do kolejnego pokolenia. Pozostałe wolne miejsca w nowej populacji są wypełniane osobnikami z najwyższą wartością v_d .

W przypadku reprodukcji ruletkowej liczność kolejnego pokolenia nie jest znana. Każdy z osobników wybierany jest bądź odrzucany z pewnym prawdopodobieństwem. W przypadku metody *dynamic sampling* liczność pochodnej populacji powinna być określona w sposób deterministyczny, tzn. liczba osobników dla których $v_d \geq 1$ jest nieznana, ale liczba "wolnego miejsca" w kolejnym pokoleniu powinna być określona.

Niech liczba miejsc w kolejnym pokoleniu będzie określona jako: $s = n_d * \alpha$, gdzie n_d to liczba osobników dla których $v_d \geq 1$, $\alpha > 1$ - stała określająca liczność kolejnego pokolenia. Wartość stałej α zostanie ustalona w sposób empiryczny.

Inną metodą określenia ilości "wolnego miejsca" może być sposób znany z metody *Stochastic Remainder Sampling* ([2]), gdzie osobniki dla których $v_d < 1$, wybierane są jak w metodzie ruletkowej.

2 Ocena metod reprodukcji

Dla pewnego problemu P (np. problem plecakowy) zostaną zaimplementowane algorytmy genetyczne uwzględniające dwie różne metody reprodukcji - metodę reprodukcji ruletkowej oraz metodę dynamic sampling.

Wyniki uruchomień obydwu algorytmów zostaną porównane przy pomocy narzędzi statystycznych. Istotność różnic średnich wartości przystosowania zostanie zweryfikowana odpowiednimi testami statystycznymi. Zestawione zostaną także uśrednione krzywe zbieżności.

Chcemy uzyskać odpowiedź na pytanie: czy zastosowanie *dynamic sampling* daje statystycznie lepsze wyniki niż zastosowanie reprodukcji ruletkowej.

3 Bibliografia

1. Arabas J., *Wykłady z algorytmów ewolucyjnych*, Wydawnictwa Naukowo-Techniczne, Warszawa, 2004.
2. Andare A.V., Errico de L., Aquino A.L.L., Assis L.P., Barbosa C.H.N.R., *Analysis of Selection and Crossover Methods used by Genetic Algorithm-based Heuristic to solve the LSP Allocation Problem in MPLS Networks under Capacity Constraints*, EngOpt 2008 - International Conference on Engineering Optimization, Rio de Janeiro, 2008.
3. Zuylen van, Anke, *Deterministic Sampling Algorithms for Network Design*, School of Operations Research and Information Engineering, Cornell University, Ithaca, NY 14853.