

Teoria algorytmów ewolucyjnych

Dlaczego teoria

- Wynik analiza teoretycznej może pokazać jakie warunki należy spełnić, aby osiągnąć zbieżność do minimum globalnego.
 - Np. sukcesja elitarystyczna.
- Może ułatwiać zrozumienia zasady działania algorytmu
 - Wiemy że działa, ale dlaczego działa.
- Może pomóc biologom teoretycznym.
- Może ułatwić zdobycie kolejnych stopni i tytułów naukowych (doktorat, habilitacja, profesura).
- Nigdy nie wiadomo gdzie może być przydatna.

Problemy z teorią

- Algorytmy ewolucyjne są bardzo skompilowanymi systemami dynamicznymi z wieloma stopni swobody.
- Typy zadań, które dobrze rozwiązują są bardzo trudne do modelowania teoretycznego.
- Duży stopień losowości oznacza:
 - konieczność użycia technik stochastycznych (np. łańcuchy Markowa).
 - wyniki analiz opisują przeciętne zachowanie.
- Po stu latach rozwoju teoretycznej biologii, naukowcy zajmujący się tą dyscypliną używają dość prymitywnych modeli prostych systemów.

Teoria schematów Johna Hollanda

- Dotyczy prostego algorytmu genetycznego (SGA):
 - reprezentacja zero-jedynkowa.
 - mutacja zmieniająca wartość bitu, krzyżowanie jednopunktowe.
 - prosta selekcja proporcjonalna bez skalowania.
- Teoria rozwijana w latach 1975-1992.
- Schemat H w $\{0,1\}^L$ jest częściowo zdefiniowanym łańcuchem zero-jedynkowym. Niezdefiniowane elementy są oznaczone jako *. (* - oznacza dowolny symbol 0 albo 1)
- Schemat definiuje hiperpłaszczyznę w $\{0,1\}^L$.
- Przykład: (L=5) (01**0).
- Zbiór wszystkich łańcuchów należących do schematu $H=(h_1, h_2, \dots, h_L)$:

$$I(H) = \left\{ (a_1, a_2, \dots, a_L) \in \{0,1\}^L : h_i \neq * \Rightarrow a_i = h_i \right\}$$

Np. $I(01**0) = \{(01000), (01010), (01100), (01110)\}$

Rząd i długość schematu

- Rząd schematu $o(H)$: liczba zdefiniowanych elementów:

$$o(H) = |\{i : h_i \in \{0,1\}\}|$$

$$o(01^{**}0) = 3$$

- Długość schematu: długość łańcucha rozpoczynającego się na pierwszym a kończącego na ostatnim zdefiniowanym elemencie.

- Np. dla $H = 1 \ 0 \ 0 \ * \ 1 \ * \ 1 \ 0 \ * \ *$ $d(H) = 7$

- Formalnie:

$$d(H) = \max \{i : h_i \in \{0,1\}\} - \min \{i : h_i \in \{0,1\}\} + 1$$

Schematy: kilka liczb

- Całkowita liczba schematów: 3^L .
- Każdy chromosom jest członkiem 2^L różnych schematów.
- W populacji o rozmiarze S reprezentowanych jest co najwyżej $S \cdot 2^L$ schematów.

Liczba kopii schematu w kolejnych iteracjach

- Zakładamy że w iteracji t znajduje się $m(H,t)$ reprezentantów schematu H .
- Przez $f(H)$ określamy średnie dopasowanie chromosomów będących reprezentantami schematu H .
- Przez \bar{f} oznaczamy średnie dopasowanie w populacji.
- Zakładając brak krzyżowania i mutacji, liczba reprezentantów schematu H w iteracji $t+1$ wynosi:

$$m(H, t+1) = m(H, t) * \frac{f(H)}{\bar{f}}$$

- Jeżeli schemat H przewyższa średnią o stałą $c > 0$, tzn. $f(H) = (\bar{f} + c * \bar{f})$, to wtedy:

$$m(H, t+1) = m(H, t) * f \frac{(H) * \bar{f} + c * \bar{f}}{\bar{f}} = m(H, t) (1 + c)$$

- Startując od $t=0$ i zakładając że c jest stałe otrzymujemy:

$$m(H, t) = m(H, 0) (1 + c)^t$$

- Wniosek liczba reprezentantów schematu o ponadprzeciętnym dopasowaniu rośnie wykładniczo.

Przeżycie schematu w procesie krzyżowania i mutacji

- Niech chromosom jest reprezentacją pewnego schematu H .
- Schemat reprezentowany przez ten chromosom ulega zniszczeniu w wyniku mutacji, jeżeli w chromosomie zostanie zmutowany bit odpowiadający 0 lub 1 w schemacie. Szansa zachowania takiego bitu wynosi $1-p_m$, gdzie p_m jest prawdopodobieństwem mutacji. Ponieważ mutacje poszczególnych alleli są od siebie niezależne, to prawdopodobieństwo przeżycia schematu w wyniku mutacji jest równe (przybliżenie prawdziwe dla małych p_m):

$$p_{sm}(H) \geq (1-p_m)^{o(H)} \approx 1 - o(H) * p_m$$

- Schemat H ulega zniszczeniu w wyniku krzyżowania jednopunktowego, jeżeli punkt krzyżowania znajdzie się pomiędzy „końcami” schematu. Przeżywa gdy punkt wypadnie poza wnętrzem schematu. Przeżycie ma miejsce z prawdopodobieństwem:

$$p_{sx}(H) \geq 1 - \frac{d(H) * p_c}{L-1}$$

- Uwaga: im krótszy schemat tym większa szansa przeżycia w wyniku krzyżowania.

Łączny wpływ trzech kroków

$$m(H, t+1) \geq m(H, t) * \frac{f(H)}{\bar{f}} * \left[1 - \frac{d(H) * p_c}{L-1} \right] * [1 - o(H) * p_m]$$

- Zależność ta nazywana jest twierdzeniem o schematach (ang. schema theorem) (Holland, 1975).
- Wniosek z twierdzenia o schematach (D.E. Goldberg, 1989):
„Schematy niskiego rzędu o małej długości, o ponad przeciętnym dopasowaniu rozprzestrzeniają się w kolejnych populacjach zgodnie z prawem wzrostu wykładniczego”

Hipoteza budujących bloków (ang. building block) (Holland, 1975; Goldberg, 1989)

- Algorytm genetyczny jest w stanie wykryć krótkie schematy o niewielkim rzędzie oraz ponadprzeciętnym dopasowaniu i stworzyć z nich wysoce dopasowanych osobników.
- Budujące bloki to schematy:
 - o małym rzędzie.
 - o niewielkiej długości.
 - o ponadprzeciętnym dopasowaniu.

Krytycyzm teorii schematów

- Jest to aproksymacja, prawdziwa dla olbrzymich rozmiarów populacji.
- W populacji o praktycznych rozmiarach, wykładniczy wzrost instancji schematu prowadzi szybko do całkowitego zapełnienia populacji.
- W populacji o praktycznych rozmiarach, wykładniczy spadek instancji schematu prowadzi szybko do jego całkowitej eliminacji.
- W populacji o praktycznych rozmiarach nie wszystkie schematy są reprezentowane.
- Średnie dopasowanie w populacji oraz średnie dopasowanie wszystkich instancji schematu nie jest stałe (zmienia się w kolejnych generacjach).
- W konsekwencji, funkcje testowe zaproponowane jako teoretycznie łatwe, w oparciu o hipotezę budujących bloków (royal road problems), okazały się lepiej rozwiązywalne przy pomocy losowego przeszukiwania lokalnego.

Analiza przy pomocy łańcuchów Markowa

- System jest łańcuchem Markowa, jeżeli:
 - Znajduje się w jednym ze skończonej liczby N stanów.
 - Prawdopodobieństwo znalezienia się w stanie X w czasie (czas jest dyskretny) $t+1$ zależy wyłącznie od stanu systemu w czasie t i nie jest zależne od stanu w czasie $t-1, t-2, \dots, 0$.
 - Drugi warunek sprawia, że możemy zdefiniować macierz przejść Q , której element $q_{i,j}$ oznacza prawdopodobieństwo przejścia ze stanu i do stanu j , przy zmianie czasu $t \Rightarrow t+1$.
- Dobrze znana teoria zachowania łańcuchów Markowa.
- Algorytm genetyczny możemy traktować jako łańcuch Markowa, w którym zbiór stanów to zbiór wszystkich możliwych populacji.
 - Ale macierze przejścia są olbrzymie, w ogólności dla algorytmu o populacji S i chromosomach zero-jedynkowych o długości L całkowita liczba stanów wynosi:

$$\begin{pmatrix} S + 2^L - 1 \\ 2^L - 1 \end{pmatrix}$$

Wynik analizy:

- Eiben, 1989. Algorytm genetyczny zbiega do optimum z prawdopodobieństwem 1 przy następujących założeniach:
 - Każdy osobnik ma niezerowe prawdopodobieństwo bycia wybranym jako rodzic.
 - Sukcesja jest elitarystyczna.
 - Każde rozwiązanie może być stworzone przy pomocy operatorów krzyżowania i mutacji, z prawd. > 0 .
- Co oznacza zbieżność z prawdopodobieństwem ?
 - Oznaczmy przez $S(t)$ populację w iteracji t . x^* rozwiązanie optymalne.
 - Wtedy:
$$\lim_{t \rightarrow \infty} P(x^* \in S(t)) = 1$$
- Wady metody:
 - nic nie mówi o sytuacji w generacji n .

Teoria algorytmów z reprezentacją zmiennopozycyjną (strategii ewolucyjnych)

- W porównaniu z algorytmami genetycznymi, teoria o wiele bardziej zaawansowana.
- Dowody zbieżności z prawdopodobieństwem.
 - w przypadku sukcesji elitarystycznej.
- Większość analiz dotyczy prostych funkcji, np. sfera $\sum_{i=1}^L x_i^2$
 - dla tych funkcji algorytm ewolucyjny nie jest najlepszym algorytmem do optymalizacji.
- Wyniki dotyczące szybkości postępu (zmiany odległości centrum masy populacji od globalnego optimum)

Twierdzenia „no free lunch” (Wolpert, Macready, 1997)

- Założenia:
 - algorytm nie odwiedza ponownie tych samych punktów w przestrzeni poszukiwań.
 - algorytm typu „black-box”: nie wykorzystuje informacji o problemie.
- Jeżeli uśrednimy po wszystkich możliwych problemach, wszystkie algorytmy optymalizacji spełniające założenia osiągają takie same wyniki.
- Wnioski:
 - Jeżeli wymyśliliśmy „cudowny” algorytm ewolucyjny działający bardzo dobrze dla wybranych problemów testowych, to na pewno będzie działał bardzo źle dla pewnych innych problemów testowych.
 - Opłaca się włączać wiedzę o problemie do algorytmów ewolucyjnych (algorytmy memetyczne).
- Do dyskusji: Czy problemy badane w praktyce są reprezentatywne dla przestrzeni wszystkich problemów.

Literatura

- Goldberg, D. E.: Algorytmy genetyczne i ich zastosowania. WNT, Warszawa, 1995, rozdział referuje teorię schematów.
- A. E. Eiben, J. Smith, Introduction to Evolutionary Computing, Springer, 2003, rozdział 11 poświęcony teorii algorytmów ewolucyjnych.