I Krajowa Konferencja
Algorytmy Ewolucyjne
Murzasichle, 12-15.06.1996

Uwagi do twierdzenia Hollanda o schematach.

Mariusz Podsiadło, Politechnika Rzeszowska, Katedra Automatyki i Informatyki
ul. W. Pola 2, 35–959 Rzeszów, e–mail: podsiadl@prz.rzeszow.pl

1. Wstęp.

Algorytmy genetyczne (GA – genetic algorithms) są od wielu lat z dużym powodzeniem stosowane do rozwiązywania wielu problemów z różnych dziedzin nauki i techniki, takich jak np. optymalizacja, badania operacyjne, genetyka, sztuczna inteligencja, przetwarzanie obrazów czy nauki społeczne (zestawienie zastosowań można znaleźć w [1]). Przekonanie o efektywności i sile GA wynika głównie z doświadczeń symulacyjnych, choć wielu badaczy prowadzi prace nad teoretycznym uzasadnieniem poprawności algorytmu. Pierwsze prace na ten temat podjął Holland. Wprowadził on pojęcie schematu i sformułował tzw. "twierdzenie o schematach" nazywane również "podstawowym twierdzeniem algorytmów genetycznych". W połączeniu z tzw. "hipotezą cegiełek" stanowi ono podstawowy aparat, przy pomocy którego tłumaczy się mechanizmy działania i przyczyny efektywności GA. Holland podjął także próbę ilościowego oszacowania efektywności GA. Podał on mianowicie oszacowanie liczby schematów, które są przez algorytm przetwarzane efektywnie.

Z punktu widzenia zastosowań GA, efektywność wydaje się być najważniejszym czynnikiem warunkującym jego użyteczność. Bliższa analiza wyników otrzymanych przez Hollanda nasuwa pewne wątpliwości. Niniejszy artykuł ma na celu przedstawienie tych wątpliwości oraz wynikających z nich wniosków.

2. Twierdzenie o schematach.

W niniejszym paragrafie zostały przypomniane podstawowe pojęcia dotyczące algorytmu genetycznego. Omówione zostały również stosowane w tekście oznaczenia. Przytoczono twierdzenie o schematach.

Algorytm genetyczny operuje na populacji (zbiorze) chromosomów (ciągów kodowych). Chromosomy są ciągami genów (alleli) mogących przyjmować wartości oznaczone symbolami określonego alfabetu. Bez utraty ogólności rozważań przyjmiemy alfabet dwójkowy $V=\{0,1\}$. Chromosomy oznaczane będą wielkimi literami alfabetu łacińskiego z ewentualnym dolnym indeksem, zaś ich elementy odpowiednimi małymi literami z dolnymi indeksami określającymi ich pozycję w ciągu. Przykładowo ciąg:

$$A=10010$$

może być symbolicznie przedstawiony następująco:

$$A=a_1a_2a_3a_4a_5$$

Schemat jest wzorcem opisującym podzbiór chromosomów podobnych ze względu na ustalone pozycje. Złożony jest z symboli rozszerzonego alfabetu $V^+=\{0,1,*\}$. Symbol "*" oznacza, że na danej pozycji ciągu może występować zarówno symbol "0" jak i "1". Schematy również oznaczane będą wielkimi literami alfabetu łacińskiego z ewentualnymi indeksami dolnymi. Przykładem schematu może być następujący ciąg symboli alfabetu V^+:

H=*0*1*

Jeżeli chromosom posiada na wszystkich pozycjach odpowiadających pozycjom ustalonym (różnym od "*") schematu dokładnie takie same wartości genów jak schemat, to mówimy, że chromosom jest reprezentantem shematu. Na przykład chromosom A jest reprezentantem schematu H, ponieważ wartości genów a_2 i a_4 są równe wartościom na pozycjach ustalonych schematu H – odpowiednio h_2 i h_4.

Każdy schemat scharakteryzowany jest liczbowo przez dwa parametry: rząd i rozpiętość. Rząd schematu, oznaczany przez $o(H)$, jest równy liczbie jego ustalonych pozycji, czyli liczbie symboli różnych od "*". Przykładowo rząd schematu H wynosi 2. Rozpiętość schematu, oznaczana przez $\delta(H)$, jest odległością między dwiema skrajnymi pozycjami ustalonymi. Przykładowo rozpiętość schematu H wynosi 4–2=2.

Załóżmy, że dany jest algorytm genetyczny z prostym krzyżowaniem i mutacją. Oznaczmy przez $A(t)$ populację złożoną z chromosomów A_j, $j=1,...,n$, w chwili (pokoleniu) t. Załóżmy również, że w populacji $A(t)$ znajduje się $m=m(H,t)$ reprezentantów danego schematu H. Zajmiemy się teraz oszacowaniem, jaka jest oczekiwana liczba wystąpień schematu H w populacji $A(t+1)$.

Podczas reprodukcji chromosomy są kopiowane do nowej populacji z prawdopodobieństwem proporcjonalnym do ich przystosowania równym $p_i = f_i / \sum f_j$. Po zakończeniu tej operacji możemy oczekiwać wystąpienia $m=m(H,t+1)$ reprezentantów schematu H w populacji $t+1$. Oczekiwana liczba schematów wynosi:

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

gdzie: $f(H)$ – średnie dopasowanie wszystkich chromosomów populacji $A(t)$ będących reprezentantami schematu H,

$\bar{f}$ – średnie dopasowanie populacji $A(t)$, $\bar{f} = \frac{1}{n}\sum_{j=1}^{n} f(A_j)$,

$E[.]$ oznacza wartość oczekiwaną.

Powyższa zależność oznacza, że oczekiwana liczba wystąpień schematu H w populacji $t+1$ wzrośnie, jeżeli średnie przystosowanie wszystkich jego reprezentantów jest większe niż średnia dopasowania całej populacji. Goldberg [1] pisze: "Pod względem jakościowym wpływ reprodukcji na schematy jest oczywisty: schematy "lepsze" od przeciętnej rozprzestrzeniają się, a "gorsze" – zanikają".

Po zakończeniu reprodukcji (selekcji) nowa populacja poddawana jest krzyżowaniu. Ponieważ chcemy ustalić, jaka jest oczekiwana liczba reprezentanów schematu H w populacji $t+1$, interesuje nas jaki wpływ będzie miało krzyżowanie na schematy. Nie trudno zauważyć,

że jeżeli punkt podziału krzyżowanych chromosomów wypadnie pomiędzy skrajnymi pozycjami schematu, to zostanie on zniszczony, o ile krzyżowane chromosomy będą posiadały różne wartości genów na pozycjach ustalonych schematu. Pomijając ten szczególny przypadek, można prawdopodobieństwo zniszczenia schematu w czasie krzyżowania wyrazić jako zależność między rozpiętością schematu a długością chromosomu: $\delta(H)/(l-1)$. Po uwzględnieniu prawdopodobieństwa krzyżowania p_c, dolne oszacowanie prawdopodobieństwa przeżycia schematu można wyrazić następująco:

$$p_{sc} \geq 1 - p_c \frac{\delta(H)}{l-1}$$

Po krzyżowaniu, populacja poddawana jest mutacji. Aby schemat H przeżył mutację, muszą się zachować wszystkie jego pozycje. Ponieważ prawdopodobieństwo zmiany genu wynosi p_m, a mutacje na poszczególnych pozycjach chromosomów są statystycznie niezależne, zatem przwdopodobieństwo przeżycia mutacji wynosi:

$$p_{sm} = (1 - p_m)^{o(H)}$$

Co można dla małych wartości p_m przybliżyć wyrażeniem $1 - o(H)p_m$.

Ostatecznie, oczekiwana liczba reprezentantów schematu H w pokoleniu $A(t+1)$ spełnia nierówność:

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

Powyższe oszacowanie zostało nazwane twierdzeniem o schematach. Zostało sformułowane i udowodnione w [2], bardziej szczegółowy dowód od przytoczonego powyżej można znaleźć również w [1] i [3].

3. Liczba schematów efektywnie przetwarzanych przez GA wg Hollanda.

W niniejszym paragrafie przedstawiono oszacowanie liczby schematów, biorących efektywny udział w przetwarzaniu, otrzymane przez Hollanda, a podane w [1] i [4]. Wynika ono bezpośrednio z interpretacji twierdzenia o schematach. Oszacowanie to mówi, że w algorytmie genetycznym, operującym na populacji n chromosomów, z prostym krzyżowaniem i niewielkim tempem mutacji, w każdej iteracji efektywnie przetwarzanych jest Cn^3 schematów. Holland nazwał to zjawisko mianem ukrytej równoległości (intrinsic parallelism). Prześledźmy teraz tok rozumowania prowadzący do wyżej wymienionego wyniku, przedstawiony w [1].

Rozważmy GA działający na populacji n chromosomów o długości l, z prostym krzyżowaniem i małym tempem mutacji. Poszukujemy liczby schematów, których reprezentacja w populacji zwiększa się w pożądany wykładniczy sposób. Interesują nas zatem schematy o prawdopodobieństwie przetrwania większym lub równym od stałej p_s, czyli schematy o rozmiarze:

$$l_s < (1 - p_s)(l-1) + 1$$

przy czym rozmiar schematu jest równy δ_s+1 (rozpiętość +1).

Liczba możliwych schematów o rozmiarze nie większym od l_s, które mogą być reprezentowane przez chromosom o długości l wynosi

$$2^{(l_s-1)}(l-l_s+1)$$

Aby zapewnić niepowtarzalność schematów w populacji przyjmijmy wielkość populacji równą

$$n=2^{l_s/2}$$

Biorąc pod uwagę fakt, że liczby schematów wyrażają się współczynnikami dwumianowymi, dochodzimy do wniosku, że połowa z nich jest rzędu wyższego niż $l/2$. Zatem dolne oszacowanie liczby różnych schematów jest następujące:

$$n_s \geq n(l-l_s+1)2^{l_s-2}$$

Uwzględniając wybraną szczególną wielkość populacji otrzymujemy zależność:

$$n_s=\frac{(l-l_s+1)n^3}{4}=Cn^3$$

Powższa zależność jest przedmiotem dyskusji dalszej części artykułu.

4. Przykład.

W niniejszym rozdziale zostanie przedstawiony przykład działania algorytmu genetycznego, na podstawie którego w dalszej części dokonamy analizy zagadnień omówionych w punktach 2 i 3. Będzie nim zadanie znalezienia maksimum funkcji kwadratowej $f(x)=x^2$, w dziedzinie $x=\{0,1,...,31\}$.

Dla celów analizy wystarczające jest przyjęcie, że populacja $\boldsymbol{A}$ składa się z czterech chromosomów o długości $l=5$. Dziedzina funkcji reprezentowana jest w postaci dwójkowej liczby całkowitej. Funkcja posiada maksimum w punkcie $11111_2 = 31_{10}$ wynoszące 961.

Załóżmy, że wylosowano następującą populację początkową:

A_1=01101
A_2=10000
A_3=00101
A_4=10110

Dopasowanie poszczególnych osobników wynosi:

$f(A_1)=f(01101_2)=13^2=169$
$f(A_2)=16^2=256$
$f(A_3)=5^2=25$
$f(A_4)=22^2=484$

średnie dopasowanie populacji $\bar{f}=233.5$

Rozważmy schematy: H_1=****0 i H_2=*1***. Schemat H_1 jest reprezentowany przez chromosomy A_2 i A_4, a schemat H_2 przez chromosom A_1. Dopasowanie schematów (średnie dopasowanie ich reprezentantów) wynosi:

$f(H_1)=(484+256)/2=370$
$f(H_2)=169$

Ponieważ schematy posiadają ten sam rząd $o(H)=1$ i rozpiętość $\delta(H)=0$, zatem biorąc pod uwagę małe prawdopodobieństwo mutacji możemy pominąć wpływ krzyżowania i mutacji na rozprzestrzenianie się schematów w populacji. Zatem oczekiwana liczba wystąpień schematów w następnej populacji jest w przybliżeniu opisana równaniem (*). Na tej podstawie stwierdzamy, że schemat H_1 ma $f(H_1)/f(H_2)=2.189$ razy większe szanse przetrwania niż H_2. Co więcej, stosunek dopasowania schematu H_1 do średniego dopasowania całej populacji wynosi 1.585, a schematu H_2 tylko 0.723. Wnioskujemy stąd, że oczekiwana liczba reprezentantów schematu H_1 wzrośnie w następnym pokoleniu, a liczba reprezentantów schematu H_2 zmaleje.

Porównajmy analizowane schematy z poszukiwanym rozwiązaniem. Łatwo zauważyć, że rozwiązanie jest reprezentantem zanikającego schematu H_2, natomiast nie jest reprezentantem rozprzestrzeniającego się schematu H_1.

5. Dyskusja.

Przykład przedstawiony w p.4 miał zilustrować źródło wątpliwości dotyczących poprawności rozważań Hollanda na temat efektywności przetwarzania GA i tzw. ukrytej równoległości. Istotnie, jeżeli przyjrzymy się schematom H_1 i H_2 pod kątem twierdzenia o schematach i rozważań z p.3, to dochodzimy do wniosku, że wzrost oczekiwanej liczby reprezentantów H_1 i spadek liczby reprezentantów H_2 jest zgodny z oczekiwaniami. Jeśli jednak wziąć dodatkowo pod uwagę rozwiązanie problemu i jego relacje ze schematami, stwierdzamy, że żaden z rozpatrywanych schematów nie może być uznany za przetwarzany efektywnie. A to jest niezgodne z tezami z p.3.

Skąd bierze się ta niezgodność? Otóż Holland formułując swoje wnioski na temat efekftywności przetwarzania schematów przez GA, niejawnie założył, że schemat jest efektywnie przetwarzany wtedy, gdy oczekiwana liczba jego reprezentantów w następnym pokoleniu rośnie wykładniczo, zgodnie z twierdzeniem o schematach. Czy jest to słuszne założenie? Nie, ponieważ wzrost liczby reprezentantów schematu danego typu, nie musi być bezpośrednio związany z jego relacją do poszukiwanego rozwiązania problemu. Innymi słowy, spodziewany wzrost lub spadek oczekiwanej liczby wystąpień danego schematu w następnej populacji nie zależy ściśle od poszukiwanego rozwiązania problemu. Aby to przeanalizować bliżej wprowadzimy na użytek niniejszego opracowania następujące definicje efektywności przetwarzania schematów.

1. Schemat jest przetwarzany efektywnie, jeżeli:

   rozwiązanie jest reprezentantem schematu i spodziewana liczba reprezentantów schematu wzrośnie w następnym pokoleniu

lub

rozwiązanie nie jest reprezentantem schematu i spodziewana liczba reprezentantów schematu zmaleje w następnym pokoleniu

2. Schemat jest przetwarzany nieefektywnie, jeżeli:
   rozwiązanie jest reprezentantem schematu i spodziewana liczba reprezentantów schematu zmaleje w następnym pokoleniu

   lub

   rozwiązanie nie jest reprezentantem schematu i spodziewana liczba reprezentantów schematu wzrośnie w następnym pokoleniu

W świetle przyjętych definicji rozpatrywane schematy H_1 i H_2 nie są przetwarzane efektywnie. Przykładem schematu przetwarzanego efektywnie może być H_3=1****. Posiada on podobnie jak H_1, dwóch reprezentantów A_2 i A_4, a zatem również takie samo dopasowanie i prawdopodobieństwo przetrwania (czyli oczekiwany wzrost liczby wystąpień w populacji).

Mamy więc dwa schematy o jednakowych parametrach (rzędzie i rozpiętości) i oczekiwanym wzroście liczby wystąpień w następnej populacji ale o diametralnie różnej efektywności ich przetwarzania. Można stąd wysnuć wniosek, iż średnie dopasowanie schematu (a więc wzrost lub spadek liczby wystąpień schematu) nie zawsze ma związek z ich znaczeniem dla poszukiwanego przez GA rozwiązania.

Można wskazać jeszcze inne schematy, które będą przetwarzane nieefektywnie:

H_4=*0***, reprezentanci A_2, A_3, A_4, $f(H_4)$=255, $f(H_4)/\bar{f}$=1.092,

H_5=**1**, reprezentanci A_1, A_3, A_4, $f(H_5)$=196.7, $f(H_5)/\bar{f}$=0.842,

oraz efektywnie:

H_6=***1*, reprezentant A_4, $f(H_4)$=484, $f(H_6)/\bar{f}$=2.073.

Porównajmy teraz dwa schematy, które uznaliśmy za przetwarzane efektywnie H_3 i H_6. Łatwo zauważyć, że dla rozwiązania problemu większe znaczenie ma schemat H_3 niż H_6. Ma on jednak mniejszą oczekiwaną liczbę wystąpień w następnej populacji, jest więc przetwarzany "mniej efektywnie" niż H_6. Wynika stąd, że wprowadzone powyżej definicje efektywności przetwarzania schematów nie uwzględniają wszystkich czynników na nią wpływających. Definicje należałoby rozszerzyć tak, aby uwzględnić właściwości funkcji dopasowania i sposób zakodowania jej dziedziny w chromosomie. Dla funkcji nieliniowych, które są najczęściej przetwarzane przez GA, nie jest to łatwe, jeżeli w ogóle możliwe. Efektywność przetwarzania schematów zależy zatem od samego zadania i sposobu aplikacji algorytmu genetycznego do jego rozwiązania. Nie można więc podać uniwersalnej zależności opisującej ilość efektywnie przetwarzanych schematów.

Nasuwa się pytanie, jak to się dzieje, że część schematów gorzej pasujących do rozwiązania jest lepiej propagowanych do następnej populacji niż schematy lepiej pasujące i na odwrót?

Przyjrzyjmy się ponownie schematom H_1 i H_4. Stwierdziliśmy wyżej, że nie są one przez algorytm przetwarzane efektywnie, gdyż nie są reprezentowane przez rozwiązanie i

posiadają przystosowanie wyższe od średniego dla całej populacji. Popatrzmy teraz na reprezentantów obu schematów. Oba schematy reprezentowane są przez chromosomy A_2 i A_4. Chromosomy te są również reprezentantami schematu H_3=1****, dzięki któremu uzyskują wysoką wartość dopasowania. Popatrzmy teraz na schemat H_2, którego jedynym reprezentantem w danej populacji jest chromosom A_1. Zwróćmy uwagę, że na najbardziej znaczącej (w sensie postawionego problemu) pozycji tego chromosomu występuje zero, przez co chromosom, a razem z nim schemat H_2, otrzymuje niewielką wartość dopasowania.

Odpowiedź na wyżej postawione pytanie można zatem ująć następująco. Wszystkie schematy są częścią reprezentujących je chromosomów, którym przypisywana jest wartość dopasowania. Jeżeli więc chromosom otrzyma wysoką ocenę dopasowania, to ta ocena staje się udziałem wszystkich schematów, które można do niego dopasować, zarówno tych, które dobrze pasują do rozwiązania, jak i tych, które do niego nie pasują, i na odwrót. Jeżeli chromosom otrzyma niską ocenę dopasowania, to ocena ta rzutuje na wszystkie zawierające go schematy "dobre" i "złe". Wynika stąd niezbicie, że nie można w ogóle mówić, że algorytm genetyczny przetwarza schematy. On przetwarza tylko i wyłącznie chromosomy, które są nierozerwalnymi zbiorami informacji o potencjalnym rozwiązaniu zadania.

6. Wnioski.

Na podstawie powyższych rozważań można sformułować następujące wnioski:

I. Fakt wzrostu oczekiwanej liczby wystąpień określonego schematu w następnej populacji, nie świadczy o jego efektywnym przetwarzaniu przez algorytm genetyczny.

II. Co więcej, nie ma sensu mówić o efektywności przetwarzania schematów, które stanowią oderwaną część rzeczywistych danych, na których operuje algorytm.

III. GA przetwarza całe chromosomy, w których zakodowane informacje o poszczególnych cechach fenotypu, są ze sobę nierozerwalnie związane; nie można zatem mówić o istnieniu ukrytej równoległości algorytmu genetycznego i prztewarzaniu przez niego Cn^3 schematów, a jedynie o przetwarzniu n chromosomów (przy czym n oznacza liczbę osobników w populacji).

IV. Twierdzenie o schematach ujmuje jedynie statystyczny związek pomiędzy liczbą wystąpień określonego schematu w kolejnych populacjach; ukryta równoległość wynika z jego błędnej interpretacji.

V. Właściwości GA opisane w dwóch ostatnich akapitach poprzedniego rozdziału są źródłem zaniku w populacji informacji istotnych z punktu widzenia poszukiwanego rozwiązania i wzrostu ilości informacji nieistotnych, a więc są ważnym czynnikiem wpływającym na globalną efektywność algorytmu.

VI. Koncepcja schematów jest dobrym narzędziem statystycznej analizy algorytmu genetycznego i niektórych zjawisk w nim zachodzących. Wątpliwe jednak wydaje się aby można było przy jej pomocy wytłumaczyć prawdziwe przyczyny efektywności i krzepkości algorytmu.

7. Podsumowanie.

W artykule przeanalizowano twierdzenie o schematach i wynikające z niego oszacowanie Hollanda dotyczące liczby efektywnie przetwarzanych przez GA schematów. Na podstawie kontrprzykładów przedstawiono wątpliwości wiązane z tym oszacowaniem. Wykazano, że GA nie przetwarza schematów (gdyż są one jedynie abstraktem – narzędziem analizy działania algorytmu) ale stanowiące nierozerwalną całość chromosomy. Otrzymane wnioski rzutują na inne zagadnienia związane z analizą sposobu działania i źródeł efektywności algorytmu genetycznego, szczególnie na tzw. hipotezę cegiełek oraz minimalny problem zwodniczy.

Znaczenie otrzymanych wniosków wynika z dotychczasowego rozumienia działania GA opartego na tezie o istnieniu "ukrytej równoległości". Autorzy piszący o GA powszechnie używają terminu "przetwarzanie schematów" w różnych formach. Znajduje to odbicie zarówno w sposobie stawiania problemów, które rozpatrują, jak i w sposobie podejścia do ich rozwiązania. Dlatego niniejsze rozważania i ich wyniki wydają się być dość istotne.

Literatura.

1. David E. Goldberg *Genetic algorithms in optimization, search and machine learning.* Addison–Wesley Publishing Company, Inc., 1989 (tłum. polskie WNT, W–wa 1995).
2. John H. Holland *Adaptation in natural and artifical sytems.* The University of Michigan Press, 1975, 1992.
3. Zbigniew Michalewicz *Genetic algorithms + data structures = evolution programs.* Springer–Verlag Berlin Heidelberg, 1992.
4. David E. Goldberg *Optimal initial population size for binary–coded genetic algorithms* (TCGA Report No. 85001). Tuscalosa: University of Alabama, The Clearinghouse for Genetic Algorithms.