

Renata ADAMCZYK
ABG S.A.

MIARY RÓŻNORODNOŚCI POPULACJI OPARTE NA ENTROPII W PROGRAMOWANIU GENETYCZNYM

Streszczenie. Artykuł przedstawia wyniki badań różnorodności populacji w programowaniu genetycznym z reprezentacją grafową. Celem głębszego zrozumienia warunków pomyślnej ewolucji algorytmów programowania genetycznego zbadano kilka miar różnorodności związanych z pojęciem entropii. Analizę korelacji miar z jakością wyników przeprowadzono za pomocą standardowych statystyk.

Słowa kluczowe: programowanie genetyczne, entropia, różnorodność populacji

BASED ON ENTROPY POPULATION DIVERSITY IN GENETIC PROGRAMMING WITH GRAPH REPRESENTATION

Summary. This paper presents a study of population diversity in genetic programming with graph representation. Several measures of diversity based on entropy are considered to gain a deeper understanding of the conditions under which the evolution of genetic programming algorithms is successful. Standard statistics are used to analyze their correlation with performance.

Keywords: genetic programming, entropy, population diversity

1. Różnorodność populacji w literaturze

1.1. Podstawowe informacje dotyczące programowania genetycznego

Programowanie genetyczne [1, 2, 3] stanowi intensywnie rozwijany kierunek obliczeń ewolucyjnych. Głównym zamierzeniem programowania genetycznego jest umożliwienie automatycznego generowania programów komputerowych bez konieczności bezpośredniego ich tworzenia. W tym celu w pamięci komputera tworzy się populacje osobników, które w standardowej wersji opisywanej metody odpowiadają syntaktycznym drzewom progra-

mów. Na podstawie wyznaczonego przystosowania dokonywana jest selekcja otrzymanych rozwiązań. W praktyce przystosowanie najczęściej wyznaczone jest przez porównanie wyników wykonania programów, stanowiących rozwiązania badanego problemu, ze zbiorem zadanych przykładów wzorcowych. Zbiór ten określa wartości danych wejściowych oraz odpowiadające im prawidłowe wyniki, które powinny zostać osiągnięte przez ewoluujące programy. Przetwarzanie powtarzane jest iteracyjnie do momentu uzyskania rozwiązania optymalnego lub spełnienia zdefiniowanych warunków, np. wygenerowania określonej liczby generacji. Jeżeli algorytm wykonywany jest poprawnie, to wysoce prawdopodobne jest otrzymywanie w kolejnych generacjach programów coraz lepiej przystosowanych. W dalszej części artykułu pojęcie populacji będzie oznaczać zbiór programów tworzących każdą z generacji badanego problemu.

1.2. Różnorodność populacji

Różnorodność stanowi kluczowy element biologicznej teorii doboru naturalnego. Dawkins [6] zauważa, że warunkiem koniecznym powstania dowolnej cechy w drodze doboru naturalnego jest występowanie w populacji genetycznej zmienności pod kątem badanej cechy. Działanie naporu selekcyjnego powoduje zmniejszenie różnorodności i ujednoczenie populacji. W tej samej pracy autor wskazuje na możliwość utrwalenia w populacji kombinacji genowych z niskim przystosowaniem. Jest to zjawisko niekorzystne, któremu zapobiec może zachowanie różnorodności. Sytuacji takiej w programowaniu genetycznym odpowiada zbieżność populacji do lokalnego optimum. Utrzymanie właściwych proporcji między naporem selekcyjnym i różnorodnością populacji stanowi jeden z podstawowych problemów programowania genetycznego. Zbyt silny napór selekcyjny może spowodować zatrzymanie się ewolucji na lokalnym optimum, z kolei zbyt wysoka różnorodność utrudnia znalezienie rozwiązania badanego problemu.

W literaturze na temat programowania genetycznego podkreśla się ważność zachowania różnorodności w zapobieganiu przedwczesnej zbieżności do lokalnego optimum [12, 14]. Wcześniejsze badania sugerują, że sukces programowania genetycznego [13] można uzyskać wtedy, gdy zapewnimy początkową różnorodność i później pozwolimy na stratę różnorodności.

Różnorodność związana jest z wszystkimi aspektami programowania genetycznego. Standardowe podejście do różnorodności odnosi się do różnic strukturalnych, które nie gwarantują różnorodności zachowań i zazwyczaj tylko wykazują, że dwie struktury nie są identyczne. Różnorodność genotypowa mierzona jest za pomocą odległości edycyjnej między drzewami [10], liczby unikatowych poddrzew [11], inicjującego materiału genetycznego [9] lub minimalnej liczby przekształceń niezbędnych do odwzorowania na siebie mierzonych struktur. Langdon [8] argumentuje, że różnorodność genotypowa jest wystarczającą górną

granicą różnorodności. Osobniki o identycznej strukturze dadzą to samo przystosowanie. Tak więc skutkiem zmniejszenia różnorodności genotypowej będzie spadek liczby unikatowych fenotypów.

Miarę sukcesu algorytmów ewolucyjnych stanowi najczęściej przystosowanie osobników. W miarach fenotypowych porównujemy różnice w wartościach przystosowania populacji w danym czasie. Rosca [18] zdefiniował entropię przystosowania. W artykule [18] przedstawił spostrzeżenie, że populacje wydają się trzymać w lokalnych optimach, kiedy entropia nie zmienia się lub maleje monotonicznie w kolejnych generacjach. Miara entropii stanowiła przedmiot badań pracy [12]. W artykule tym wykazano wysoką korelację entropii z jakością rozwiązań kilku problemów, między innymi mrówkowego i parzystości.

Należy zaznaczyć, że w literaturze istnieje niewiele opracowań dotyczących różnorodności populacji w programowaniu genetycznym z reprezentacją grafową. Niniejszy artykuł jest próbą zapalenia tej luki.

2. Programowanie genetyczne z reprezentacją w postaci grafu skierowanego

Reprezentację chromosomu w postaci grafu skierowanego wprowadzili do programowania genetycznego Teler i Veloso [20, 21] w systemie PADO (*Parallel Algorithm Discovery and Orchestration*) oraz Poli [17] w systemie PDGP (*Parallel Distributed Genetic Programming*). Zastosowanie reprezentacji grafowej umożliwiło rozszerzenie zastosowań programowania genetycznego na nowe klasy problemów wymagające zrównoleglenia obliczeń. Dodatkową korzyść stanowiło efektywniejsze przeszukanie przestrzeni rozwiązań oraz możliwość przewyciężenia nadmiernego wzrostu rozmiaru kodu programów (ang. *bloat*) dzięki ponownemu wykorzystaniu w grafach wcześniej zdefiniowanych struktur.

W systemie PDGP programy reprezentowane są za pomocą grafów z węzłami zawierającymi funkcje i symbole terminalne oraz łukami przedstawiającymi przepływ sterowania i wyników. System ten z powodzeniem zastosowano w rozwiązaniu problemu strzyżenia trawnika, symbolicznej regresji oraz rozpoznawania języka naturalnego.

W systemie PADO wierzchołki grafu zawierają instrukcje wykonawcze i warunki przejścia do połączonych wierzchołków. System ten zastosowano z powodzeniem w problemie rozpoznawania twarzy [20] i klasyfikacji obiektów [21].

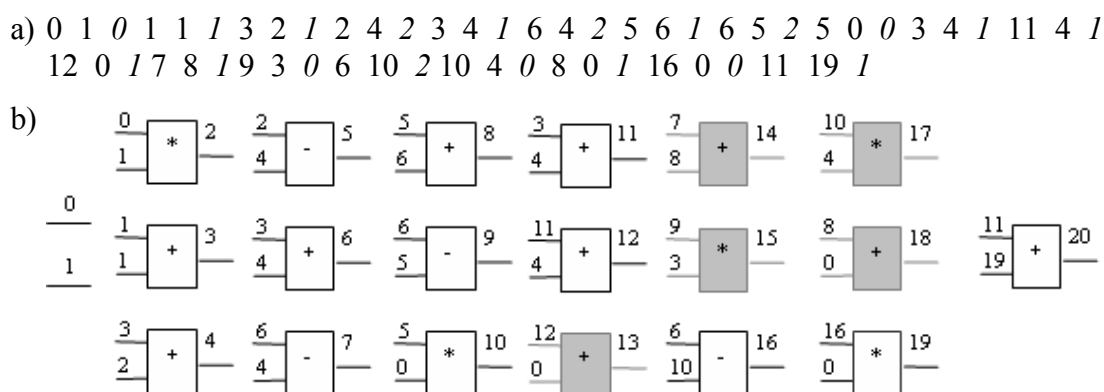
W systemach PADO i PDGP postać genotypowa poszukiwanego rozwiązania jest taka sama jak postać fenotypowa. Podejście takie związane było z dużymi trudnościami implementacyjnymi. Miller i Thomson w pracy [16] wprowadzili system CGP (*Cartesian Genetic Programming*). W proponowanym rozwiązaniu autorzy połączyli korzyści, jakie oferują

struktury grafowe z prostotą implementacji genotypu w postaci łańcuchów liczb. Odwzorowanie genotypu na fenotyp dokonywane jest za pomocą algorytmu, który przekształca łańcuchy liczb w grafy.

W systemie CGP program reprezentuje graf zakodowany w postaci tekstowych łańcuchów liczb. Symbole terminalne, funkcje oraz generowane w czasie ewolucji bloki funkcjonalne numerowane są sekwencyjnie. Genotyp odwzorowywany jest w indeksowany graf, który można wykonać jako program. Łańcuchy liczb stanowiące genotyp rozwiązania składają się z elementów nazywanych w dalszej części artykułu blokami funkcjonalnymi. Bloki funkcjonalne tworzone są w czasie działania algorytmu w następujący sposób:

- Funkcje stanowiące symbole nieterminalne rozwiązania numerowane są kolejnymi liczbami naturalnymi, począwszy od zera dla pierwszego elementu.
- Symbole terminalne numerowane są kolejnymi liczbami naturalnymi, począwszy od zera dla pierwszego elementu.
- Kolejne bloki funkcyjne numerowane są kolejnymi liczbami naturalnymi, począwszy od numeru większego o jeden od największego numeru nadanemu symbolowi terminalnemu.
- Każdemu blokowi funkcyjnemu przypisywane są: numer funkcji wykonywanej w bloku oraz numery parametrów wejściowych funkcji, wybieranych spośród numerów symboli terminalnych oraz numerów bloków funkcyjnych wcześniej wygenerowanych.

Rysunek 1 przedstawia graf systemu CGP, będący rozwiązaniem problemu regresji dla funkcji $3x^2y + 5xy + 4y$. W rozwiązaniu numery 0 i 1 odpowiadają symbolom terminalnym x i y . Operacji mnożenia odpowiada numer 0, 1 odpowiada dodawaniu, a 2 odejmowaniu. Numery funkcji zaznaczono w genotypie kursywą. Bloki funkcyjne, które nie zostały wykorzystane jako wejścia dla następujących po nich bloków, nazywamy niepołączonymi (oznaczone szarym kolorem). Bloki takie pełnią rolę intronów, czyli nieużywanych fragmentów kodu. Stanowią one dodatkowy materiał genetyczny dla ewolucji.



Rys. 1. Graf systemu CGP: a) genotyp rozwiązania problemu regresji, b) graf odwzorowujący genotyp na fenotyp

Fig. 1. The CGP's graph: a) which is a solution for a regression problem, b) which maps a genotyp to a fenotyp

System CGP zastosowano z powodzeniem w rozwiązywaniu problemów spełnialności funkcji boolowskich oraz automatycznej ewolucji obwodów elektrycznych [7].

3. Entropia

3.1. Historia pojęcia entropii oraz zastosowania w nauce

Pojęcie entropii pełni ważną rolę w wielu dziedzinach współczesnej nauki. Wykorzystywane jest w tak różnych naukach, jak fizyka, statystyka, teoria informacji, kompresja danych, lingwistyka matematyczna, nauka o fraktalach, biologia, ekonomia oraz mechanika kwantowa. Historycznie pojęcie entropii wprowadzono do nauki w wyniku badań zamiany energii cieplnej na pracę. W pierwszej połowie osiemnastego stulecia francuski fizyk Sadi Carnot wykazał, że ciepło jest w stanie wykonać pracę tylko przy przejściu z ciała cieplejszego do zimniejszego. W połowie osiemnastego wieku niemiecki uczony Rudolf Clausius ustalił wielkość równą stosunkowi wymienionego ciepła do temperatury i nazwał ją „entropią”. W 1865 r. ten sam uczony sformułował drugą zasadę termodynamiki w postaci: „Niemożliwe jest przekazywanie ciepła od ciała chłodnego do gorącego bez wystąpienia trwałych zmian w otoczeniu”. W drugiej połowie dziewiętnastego wieku Ludwik Boltzmann sformułował statystyczną interpretację drugiej zasady termodynamiki, zgodnie z którą entropia jest funkcją stanu wyrażonego liczbą możliwych sposobów jego realizacji w danej temperaturze.

Aktualnie entropia znana jest powszechnie jako miara nieporządku panującego w układzie fizycznym. Stan równowagi o najwyższej entropii jest stanem najbardziej prawdopodobnym do zaistnienia. Zmniejszenie entropii układu równoznaczne jest ze wzrostem stopnia organizacji układu. Procesy w układach fizycznych bez oddziaływań zewnętrznych rozwijają się w kierunku wzrostu entropii.

W matematyce entropia jest miarą nieokreśloności przewidywania. W przypadku zbioru zdarzeń losowych z przypisanym prawdopodobieństwem wystąpienia podaje dokładność przewidzenia następnego zdarzenia. Entropia osiąga wartość maksymalną dla równych prawdopodobieństw wystąpienia wszystkich zdarzeń. W miarę różnicowania się prawdopodobieństw entropia maleje i osiąga wartość minimalną dla zdarzenia pewnego. Wzrost entropii można również wywołać zwiększając liczbę zdarzeń losowych.

Do teorii informacji pojęcie entropii wprowadził Shannon [4], definiując ją jako miarę nieokreśloności informacyjnej układu. Entropia równa jest średniej ilości informacji przypadającej na pojedynczą wiadomość. Zmniejszenie entropii oznacza w tym kontekście spadek nieokreśloności, czyli wzrost pewności przewidywania. Możemy przyjąć, że ubytek entropii równy jest przyrostowi ilości informacji.

W praktyce istnieją procesy zmniejszania entropii, czyli zwiększania porządku. Są to np. procesy gospodarcze mające na celu koncentrację energii oraz wytwarzanie produktów. Spadek entropii jest jednak tylko lokalny, ponieważ globalna entropia rośnie wskutek zużycia energii. Procesy te stanowią przedmiot badań ekonomii i nauk o zarządzaniu.

Uczeni zauważyli sprzeczność drugiej zasady termodynamiki z ewolucją organizmów żywych. Proces ewolucji przebiega od układów prostych do złożonych. W naturze stopniowo wyłaniają się formy o coraz większym stopniu skomplikowania i organizacji, mimo że zgodnie z drugą zasadą termodynamiki powinny podlegać niepolepszeniu, a rozkładowi. Istnieje wiele prób wytłumaczenia tego paradoksu, między innymi stwierdzenie, że organizmy stanowią systemy otwarte i jako takie czerpią energię z otoczenia obniżając entropię własną. Wzrastająca entropia wewnętrzna organizmów jest przyczyną ich śmierci. Spadek entropii organizmu lub układu możliwy jest kosztem energii dostarczonej z otoczenia.

3.2. Miary różnorodności związane z pojęciem entropii

3.2.1. Znaczenie miar opartych na entropii w programowaniu genetycznym

Artykuł przedstawia propozycję wprowadzenia do programowania genetycznego CGP miar różnorodności związanych z pojęciem entropii. Oprócz miary entropii przystosowania wprowadzonej przez Rosca [18] zaproponowano miarę entropii rozproszenia bloków funkcyjnych oraz kilka miar badających wzajemną korelację badanych entropii. Idea badań wywodzi się z obserwacji, że wprowadzenie wzajemnych powiązań między elementami układu powoduje zmniejszenie jego entropii. Prawdopodobieństwo przewidzenia następnego stanu układu ulega wówczas zwiększeniu. Można więc sterować stanem układu, zmieniając zależności między jego elementami. Znajomość zależności między przystosowaniem i występowaniem bloków funkcyjnych może w konsekwencji umożliwić opracowanie skutecznej metody kontroli przebiegu procesu ewolucji. W przypadku wykrycia zbieżności do optimum lokalnego, korzystne może okazać się zwiększenie wybranych rodzajów entropii przez wpływ na wielkości z nimi skorelowane. Dodatkowa entropia może wprowadzić do układu różnorodność pozwalającą na wyjście z optimum lokalnego. Ważne jest opracowanie metody pozwalającej kontrolować proces ewolucji oraz ingerować w niego we właściwym czasie i we właściwy sposób.

3.2.2. Definicje

Rozpatrzmy dwie zmienne losowe:

- zmienną P oznaczającą przystosowanie;
- zmienną B oznaczającą występowanie bloków funkcyjny.

Entropią przystosowania [18] nazywamy wielkość zdefiniowaną wzorem:

$$H(P) = -\sum_k p_P(k) \cdot \log p_P(k) \quad (1)$$

gdzie $p_P(k)$ oznacza prawdopodobieństwo wystąpienia przystosowania k w populacji osobników.

Entropią rozproszenia bloków funkcyjnych nazywamy wielkość zdefiniowaną wzorem:

$$H(B) = -\sum_l p_B(l) \cdot \log p_B(l) \quad (2)$$

gdzie $p_B(l)$ oznacza prawdopodobieństwo wystąpienia bloku funkcyjnego l w zbiorze wszystkich bloków funkcyjnych.

Entropia jest zawsze nieujemna. Entropia jest równa zero, gdy jedno zdarzenie występuje z prawdopodobieństwem równym jedności, a pozostałe mają prawdopodobieństwa równe zero. Wartość maksymalną osiąga w przypadku, gdy prawdopodobieństwa wszystkich zdarzeń są równe.

Rozpatrzmy zmienną losową mierzącą łączne wystąpienie wybranej wartości przystosowania oraz danego bloku funkcyjnego. **Entropią łączną** [5] (ang. *joint entropy*) zmiennych losowych P i B nazywamy wielkość:

$$H(P, B) = -\sum_k \sum_l p_{P,B}(k, l) \cdot \log p_{P,B}(k, l) \quad (3)$$

Załóżmy teraz, że po pomiarze zmiennej losowej B znajdujemy wynik l , co zmienia prawdopodobieństwo zdarzenia k z $p_P(k)$ na prawdopodobieństwo warunkowe $p_{P,B}(k|l)$.

Entropią warunkową [5] przystosowania P pod warunkiem wystąpienia bloku funkcyjnego B nazywamy wielkość:

$$H(P | B) = -\sum_l \sum_k p_{P,B}(k, l) \cdot \log p_{P,B}(k | l) \quad (4)$$

gdzie $p_{P,B}(k|l)$ oznacza prawdopodobieństwo warunkowe wystąpienia przystosowania k pod warunkiem zaistnienia bloku l .

Entropią warunkową [5] wystąpienia bloku funkcyjnego B pod warunkiem zaistnienia przystosowania P nazywamy wielkość:

$$H(B | P) = -\sum_k \sum_l p_{P,B}(k, l) \cdot \log p_{P,B}(l | k) \quad (5)$$

gdzie $p_{P,B}(l|k)$ oznacza prawdopodobieństwo warunkowe wystąpienia bloku funkcyjnego l pod warunkiem zaistnienia przystosowania k .

Informacją wzajemną [5] (ang. *mutual information*) przystosowania P i rozproszenia bloków funkcyjnych B nazywamy wielkość:

$$I(P, B) = - \sum_k \sum_l p_{P,B}(k, l) \cdot \log \frac{p_{P,B}(k, l)}{p_P(k) \cdot p_B(l)} \quad (6)$$

Informacja wzajemna jest miarą zależności między zmiennymi. Wartość zero przyjmowana jest dla zmiennych niezależnych.

Spójnością informacji [19] (ang. *informational coherence*) przystosowania P i rozproszenia bloków funkcyjnych B nazywamy wielkość:

$$IC(P, B) = \frac{I(P, B)}{\min(H(P), H(B))} \quad (7)$$

Można wykazać następujące właściwości wprowadzonych pojęć [5]:

$$I(P, B) \leq H(P) - H(P|B), \quad I(P, B) \leq H(B) - H(B|P) \quad (8)$$

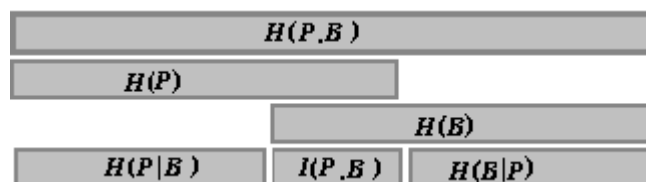
$$H(P|B) \leq H(P), \quad H(B|P) \leq H(B) \quad (9)$$

$$I(P, B) = I(B, P), \quad I(P, P) = H(P), \quad I(B, B) = H(B) \quad (10)$$

$$I(P, B) \leq H(P), \quad I(P, B) \leq H(B) \quad (11)$$

$$I(P, B) \leq 0, \quad \text{dla } P \text{ i } B \text{ niezależnych } I(P, B) = 0 \quad (12)$$

Rysunek 2 przedstawia intuicyjną reprezentację zdefiniowanych miar.



Rys. 2. Intuicyjna reprezentacja przedstawionych wartości

Fig. 2. An intuitional representation of presented values

4. Badane wielkości i problemy

W niniejszym rozdziale zostaną przedstawione wielkości, których korelacja z miarami entropii zostanie zbadana w części eksperymentalnej oraz krótki opis badanych problemów.

Przystosowanie dostosowane (ang. *adjusted fitness*) wyznaczane jest na podstawie przystosowania pierwotnego, zgodnie ze wzorem:

$$a(p) = \frac{1}{1 + f(p)} \quad (13)$$

gdzie: $a(p)$ – przystosowanie dostosowane populacji, $f(p)$ – przystosowanie pierwotne.

Dla wszystkich problemów wartości przystosowania dostosowanego dążą od zera do jedności dla coraz lepszych rozwiązań, natomiast wartości przystosowania pierwotnego dążą od

maksymalnej wartości specyficznej dla badanego problemu do zera dla coraz lepszych rozwiązań.

Problem mrówkowy stanowi próbę wyznaczenia najlepszej strategii zebrania 89 kulek żywności rozmieszczonych na dwuwymiarowej siatce 32 x 32. Zbiór funkcji F i symboli terminalnych T zdefiniowano następująco:

- $F = \{\text{if food ahead, progn2, progn3}\}$,
- $T = \{\text{left, right, move}\}$.

Przystosowanie pierwotne równe jest liczbie pominiętych kulek.

Problem regresji stanowi próbę wyznaczenia zadanego wzoru funkcji. W pracy badano funkcję o postaci: $3x^2y + 5xy + 4y$. Zbiór funkcji F i symboli terminalnych T zdefiniowano następująco: $F = \{+, -, *\}$, $T = \{x, y\}$. Przystosowanie pierwotne zdefiniowano za pomocą wzoru:

$$f(p) = \sum_{i=1}^n |p(i) - o(i)| \quad (14)$$

gdzie: $f(p)$ – przystosowanie programu, n – liczebność zbioru wzorcowych przypadków, $p(i)$ – wynik proponowanego rozwiązania programu genetycznego dla i -tego przykładu wzorcowego, $o(i)$ – prawidłowy wynik dla tego przykładu.

Problem strzyżenia trawnika stanowi próbę wyznaczenia najlepszej strategii dla kosiarki poruszającej się po planszy o wymiarach 8 x 8. Zbiór funkcji F i symboli terminalnych T zdefiniowano następująco:

- $F = \{\text{v8a, progn2, frog}\}$,
- $T = \{\text{lawnERC, left, move}\}$.

Przystosowanie pierwotne równe jest liczbie pominiętych kulek.

Korelację badanych wielkości wyznaczono za pomocą **współczynnika rho Spearmana** [23], będącego rangową miarą zależności dwóch zmiennych:

$$W_R = \frac{3 \cdot \sum_{i=1}^N ((2 \cdot \text{rank}(x_i) - N - 1) \cdot (2 \cdot \text{rank}(y_i) - N - 1))}{N(N^2 - 1)} \quad (16)$$

gdzie: N – liczba generacji, x_i – wartość pierwszej badanej zmiennej dla i -tej generacji, y_i – wartość drugiej badanej zmiennej dla i -tej generacji, $\text{rank}()$ – ranga wartości, czyli kolejność po uporządkowaniu wartości od najmniejszej do największej w generacji.

Wartości współczynnika zawierają się w przedziale domkniętym od 1 do -1. Znak współczynnika wskazuje na znak korelacji. Wartość bezwzględna współczynnika wskazuje na moc korelacji.

5. Środowisko przeprowadzenia eksperymentu

Eksperyment polegał na stukrotnym rozwiązaniu każdego z badanych problemów: strzyżenia trawnika, regresji i mrówkowego. Testy przeprowadzono w środowisku ECJ (*Evolutionary Computation in Java, Academic Free License version 3.0*) [15], którego standardową dystrybucję rozszerzono o implementację systemu CGP oraz przedstawionych miar różnorodności.

Tabela 1 przedstawia parametry wykonania algorytmu programowania genetycznego.

Tabela 1

Parametry wykonania algorytmu programowania genetycznego

Parametr	Wartość
Rozmiar populacji	1024
Inicjalizacja	Mieszana (ang. <i>ramped half-and-half</i>)
Minimalna liczba bloków funkcyjnych	4
Maksymalna liczba bloków funkcyjnych	30
Metoda selekcji	Selekcja turniejowa
Rozmiar turnieju	7
Prawdopodobieństwo operacji krzyżowania	0.6
Prawdopodobieństwo operacji mutacji	0.3
Kryterium stopu	51 generacji lub osobnik z zerowym przystosowaniem

6. Wyniki eksperymentu

Wyniki eksperymentu uporządkowano w kolejności od najlepszych do najgorszych. Na pierwszym miejscu umieszczono wykonanie z najlepszym rozwiązaniem otrzymanym w najmniejszej liczbie generacji, z najlepszym przystosowaniem średnim populacji. Kolejne miejsca zajmują wykonanie z gorszym przystosowaniem i dłuższym czasem generacji.

Dla wykresów korelacji wykreślono linię trendu. Na wykresach problemów strzyżenia trawnika i regresji oddzielono szarą pionową linią rozwiązania optymalne od pozostałych. Dla problemu mrówkowego nie uzyskano rozwiązań optymalnych.

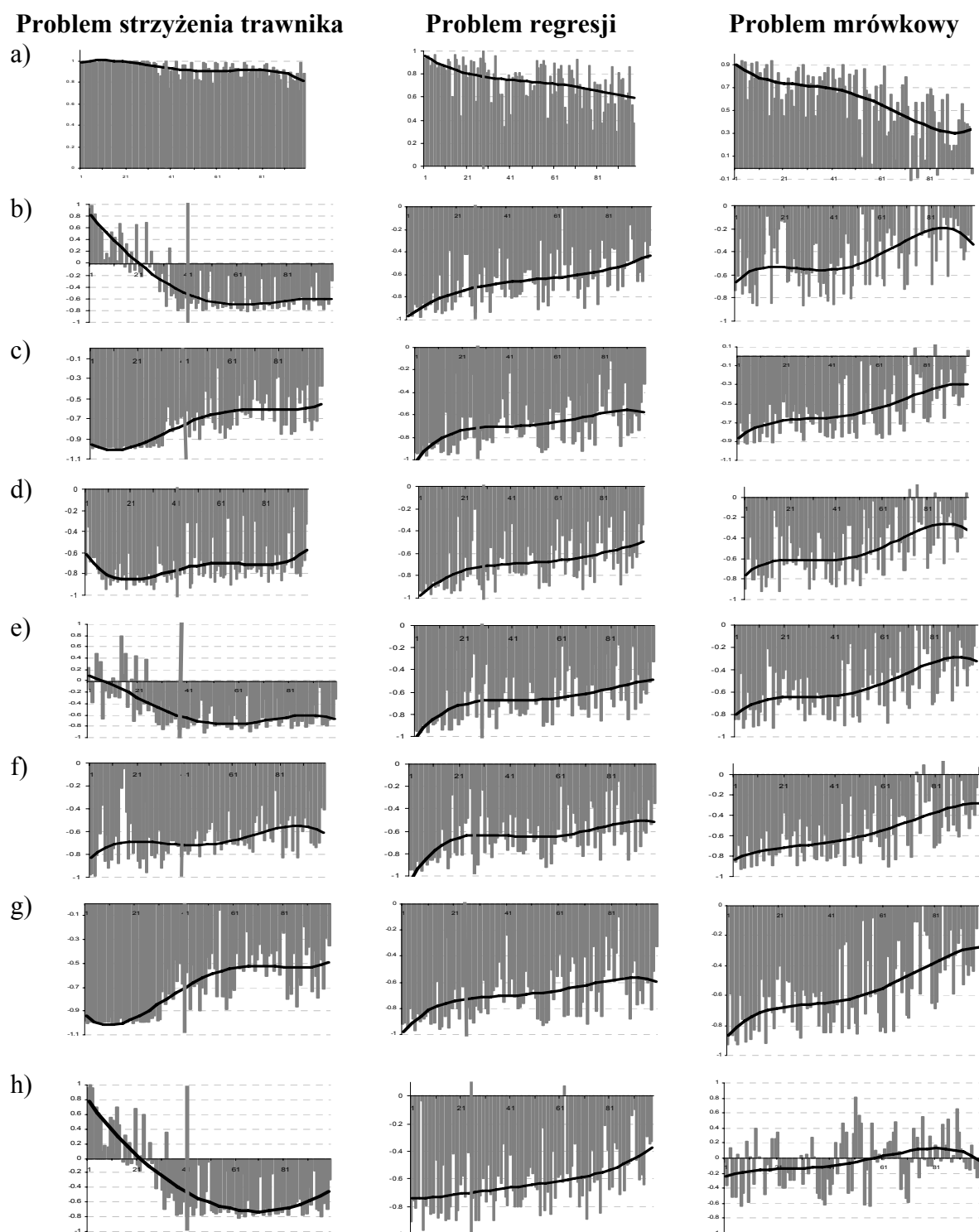
6.1. Analiza korelacji przystosowania z badanymi miarami

W artykule zaprezentowano wyniki korelacji przystosowania najlepszego osobnika. Korelacje przystosowania średniego przedstawiają się analogicznie i w związku z tym zostały pominięte. Na osi odciętych układu współrzędnych zaznaczono kolejne wykonania uporządkowane w kolejności od najlepszych do najgorszych. Na pierwszym miejscu umieszczono wykonanie z najlepszym rozwiązaniem otrzymanym w najmniejszej liczbie generacji, z naj-

lepszym przystosowaniem średnim populacji. Kolejne miejsca zajmują wykonanie z gorszym przystosowaniem i dłuższym czasem generacji, aż do wykonań najslabszych, czyli takich, w których otrzymano najgorsze rozwiązania w największej liczbie generacji. Oś rzędnych przedstawia wartości współczynników korelacji rho Spearmana. Wartości współczynników korelacji zaprezentowano na wykresach kolumnowych z wykreśloną linią trendu. Linia trendu obrazuje zmienność współczynnika korelacji dla wykonań najlepszych do najslabszych, wskazując, które korelacje w procesie ewolucji są najkorzystniejsze pod względem otrzymania optymalnych rozwiązań.

Analiza wykresów przedstawionych na rys. 3 pozwala na wyciągnięcie następujących wniosków, dotyczących zmienności wartości współczynnika korelacji rho Supermana dla wykonań uporządkowanych według jakości otrzymanych rozwiązań:

- Współczynnik korelacji ze średnim przystosowaniem generacji (rys. 3a) osiąga wysoką dodatnią wartość dla najlepszych wykonań wszystkich badanych problemów. Dla gorszych wykonań współczynnik korelacji przyjmuje coraz mniejsze wartości.
- Współczynnik korelacji z miarą entropii przystosowania dla najlepszych wykonań problemu strzyżenia trawnika (rys. 3b) przyjmuje wysoką dodatnią wartość, natomiast dla problemów regresji i mrówkowego wysoką wartość ujemną. Dla słabszych generacji wszystkich badanych problemów istnieje malejąca ujemna korelacja.
- Współczynnik korelacji z miarą entropii rozproszenia bloków funkcyjnych (rys. 3c) przyjmuje wysoką ujemną wartość dla najlepszych wykonań wszystkich problemów. Dla słabszych wykonań wszystkich problemów korelacja staje się coraz mniejsza, aż do braku korelacji w problemie mrówkowym.
- Współczynnik korelacji z entropią łączną (rys. 3d) wykazuje podobne zmiany wartości jak współczynnik korelacji z entropią blokową, z tą różnicą, że dla gorszych wykonań jest mniejsza. Dla najlepszych wykonań problemu strzyżenia trawnika korelacja jest nieco słabsza, co może odzwierciedlać różnice w korelacji dla entropii przystosowania tego problemu.
- Współczynnik korelacji z informacją wzajemną (rys. 3e) osiąga wysoką wartość ujemną dla najlepszych wykonań problemu regresji i mrówkowego. Dla analogicznych wykonań problemu strzyżenia trawnika widoczny jest brak korelacji. Dla słabszych wykonań wszystkich badanych problemów współczynnik korelacji osiąga niewielką wartość ujemną.
- Współczynnik korelacji ze spójnością informacji (rys. 3f) osiąga wysoką wartość ujemną dla dobrych wykonań wszystkich badanych problemów. Dla słabszych wykonań wszystkich problemów współczynnik korelacji przyjmuje niskie wartości.



Rys. 3. Korelacje najlepszego przystosowania dostosowanego z: a) przystosowaniem średnim, b) entropią przystosowania, c) entropią blokową, d) entropią łączną, f) informacją wspólną, g) entropią względną bloku względem przystosowania, h) entropią względną przystosowania względem bloku

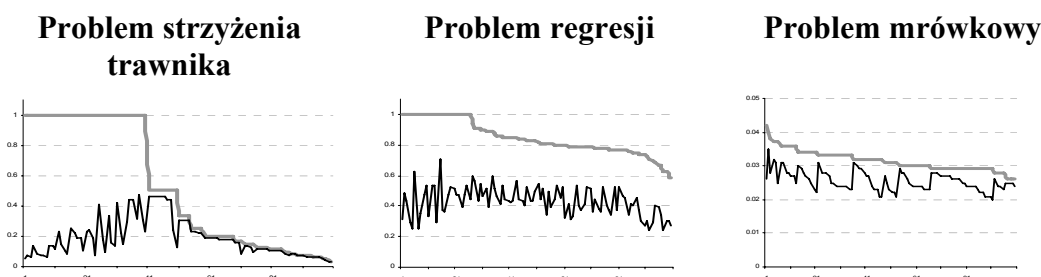
Fig. 3. Correlations between best adjusted fitness and: a) average fitness, b) fitness's entropy, c) block's entropy, d) joint entropy, e. mutual information, f) informational coherence, g) relative entropy blocks and fitness, h) relative entropy fitness and blocks

- Współczynnik korelacji z entropią warunkową rozproszenia bloków względem przystosowania dla wszystkich badanych problemów (rys. 3g) osiąga wysoką wartość ujemną dla dobrych wykonań wszystkich badanych problemów. Dla wykonań gorszych korelacja staje się coraz słabsza.
- Współczynnik korelacji z entropią warunkową przystosowania względem rozproszenia bloków dla najlepszych wykonań problemu strzyżenia trawnika (rys. 3h) przyjmuje wysoką dodatnią wartość, natomiast dla problemu regresji ujemną. Dla słabszych wykonań obu problemów współczynnik korelacji przyjmuje niewielką wartość ujemną. Dla problemu mrówkowego współczynnik korelacji przyjmuje wartości bliskie zeru, co oznacza, że korelacja praktycznie nie występuje.

Przeprowadzona analiza wyników wykazuje, że dla wszystkich badanych problemów podobne korelacje zachodzą dla następujących miar różnorodności: entropii rozproszenia bloków funkcyjnych, entropii łącznej, spójności informacji, entropii warunkowej rozproszenia bloków względem przystosowania. Wyniki eksperymentu pozwoliły więc na wskazanie miar, dla których wartości współczynnika korelacji rho Spearmana wykazują podobny przebieg zmienności dla wszystkich badanych problemów. Dla każdej z tych miar warto rozważyć możliwość wykorzystania w metodzie kontroli przebiegu ewolucji. Metoda taka mogłaby być niezależna od badanego problemu.

6.2. Analiza wartości przystosowania ostatnich generacji

Rysunek 4 przedstawia analizę zmienności przystosowań ostatnich generacji badanych wykonań. Kolorem szarym zaznaczono najlepsze przystosowanie, a kolorem czarnym przystosowanie średnie. Wykonania uporządkowano od najlepszego do najslabszego. Na osi odciętych układu współrzędnych zaznaczono kolejne wykonania. Oś rzędnych przedstawia wartości przystosowania.



Rys. 4. Wartości najlepszego i średniego przystosowania

Fig. 4. Best and raw fitness' values

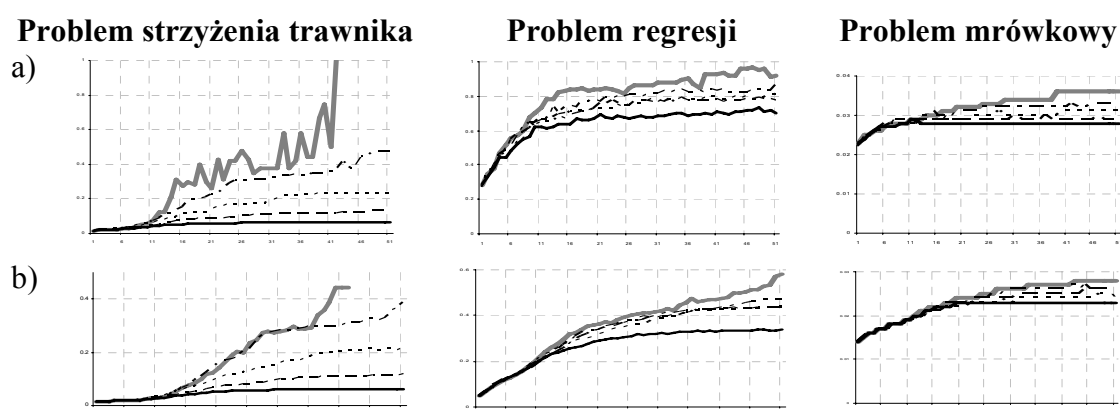
Wartości przystosowania ostatnich generacji badanych wykonań najkorzystniejszych rozwiązań znacznie różnią się od wartości przystosowania średniego. Dla słabszych generacji różnica staje się coraz mniejsza. W przypadku problemu strzyżenia trawnika różnica ta praktycznie zanika.

Porównując wyniki ze współczynnikami korelacji analizowanymi w poprzednim rozdziale, zauważamy, że silnej korelacji towarzyszy duża różnica wartości przystosowania najlepszego i średniego. Słaba korelacja związana jest z małą różnicą wartości. Obserwacja ta może wskazywać, że dla słabych generacji zakończyła się faza zbieżności do najlepszego rozwiązania, w populacji pojawiają się osobniki o różnym przystosowaniu, ale nie następuje już zbieżność do optymalnego rozwiązania.

6.3. Analiza zmian przystosowania w procesie ewolucji

Rysunek 5 przedstawia analizę zmian przystosowania najlepszego i średniego w procesie ewolucji. Na osi odciętych układu współrzędnych zaznaczono kolejne generacje. Oś rzędnych przedstawia wartości przystosowania. W poprzednim podrozdziale analizowano zmienność przystosowania między ostatnimi generacjami wykonań. W niniejszym punkcie analizujemy zmiany przystosowania w czasie wykonania.

Wszystkie wykonania podzielono na pięć grup zgodnie z kolejnością osiąganych rezultatów, od najlepszych do najslabszych. Dla każdej grupy obliczono średnie przystosowanie najlepsze i średnie dla każdej generacji. Wyniki obliczeń przedstawiono na wykresie. Szara ciągła linia odpowiada najlepszej grupie, skupiającej wykonania z najlepszymi wynikami. Czarna ciągła linia przedstawia zmiany wartości najslabszych wykonań. Linie kreskowane reprezentują grupy z wynikami zawartymi między najlepszym i najslabszym rozwiązaniem.



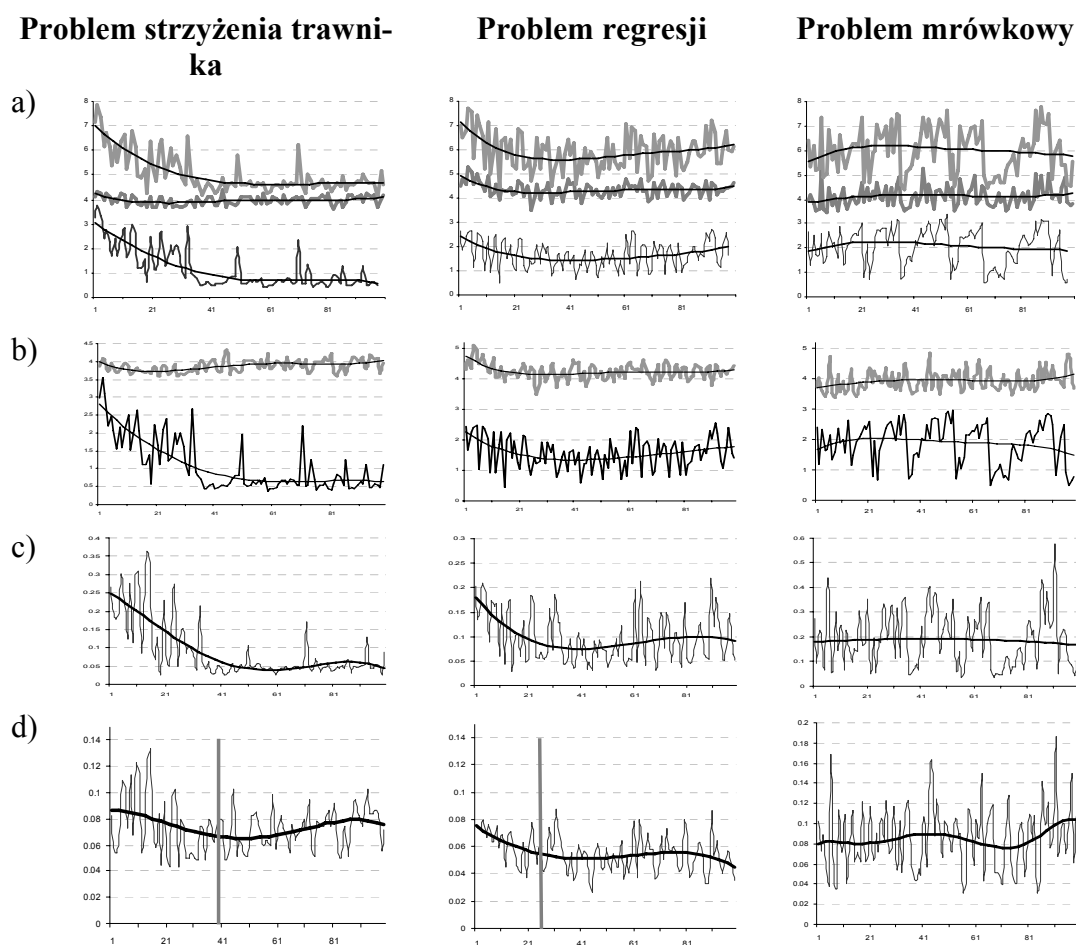
Rys. 5. Wartości przystosowania dla generacji: a) najlepszego, b) średniego
Fig. 5. Fitness' values for generations: a) best, b) average

Analiza zmian wartości najlepszego przystosowania (rys. 5a) wskazuje, że wartości przystosowań grup najlepszych i najslabszych praktycznie nie różnią się do około piątej generacji problemu regresji i strzyżenia trawnika oraz dziesiątej generacji problemu mrówkowego. Dla kolejnych generacji wartości przystosowania szybko rosną dla korzystnych wykonań, natomiast dla wykonań słabych wartości przystosowania do około dwudziestej generacji praktycznie nie zmieniają się. Zmiany średniego przystosowania (rys. 5b) przedstawiają się analogicznie, z tą różnicą że zmiany wartości są łagodniejsze.

Zmiany przystosowań wszystkich badanych problemów wykazują więc podobną zmienność dla wykonań od najlepszych do najslabszych.

6.4. Analiza wartości miar entropii ostatnich wykonań

Rysunek 6 przedstawia analizę zmienności miar entropii ostatnich generacji badanych wykonań. Wykonania uporządkowano od najlepszego do najslabszego. Na osi odciętych układu współrzędnych zaznaczono kolejne wykonania. Oś rzędnych przedstawia wartości miar entropii.



Rys. 6. Wartości miar: a) entropii przystosowania, blokowej i łącznej, b) warunkowej, c) informacji wspólnej, d) spójności informacji

Fig. 6. Measures' values: a) fitness, block and joint entropy, b) relative entropy, c) mutual informational, d) informational coherence

Analiza wykresów przedstawionych na rys. 6 pozwala na wyciągnięcie następujących wniosków:

- Górna linia wykresów (rys. 6a) przedstawia entropię łączną, środkowa entropią blokową, a najniższa entropię przystosowania. Porównanie wyników wykazuje, że dla problemu strzyżenia trawnika dla najlepszych wykonań wartości entropii łącznej i przystosowania

są znacznie wyższe od wartości dla najslabszych wykonań. Dla problemu regresji wartości entropii tych wykonań są nieznacznie większe, a dla problemu mrówkowego praktycznie nie zmieniają się. Wartość entropii blokowej jest praktycznie taka sama dla wszystkich wykonań, tylko dla problemu strzyżenia trawnika i regresji jest nieznacznie wyższa dla najlepszych wykonań.

- Górna linia wykresów (rys. 6b) przedstawia entropię warunkową wystąpienia bloku względem przystosowania, dolna linia przedstawia entropię warunkową przystosowania względem wystąpienia bloku. Porównanie wartości wskazuje na wyższe wartości entropii warunkowych dla wykonań korzystnych problemu strzyżenia trawnika i regresji. Dla problemu mrówkowego zauważalna jest większa zmienność entropii warunkowej przystosowania względem bloku, bez określonej prawidłowości.
- Wartości miary informacji wzajemnej (rys. 6c) wykazują wyższe wartości dla wykonań korzystnych problemu strzyżenia trawnika i regresji, dla słabszych wykonań problemu mrówkowego zauważalna jest większa zmienność miary, bez określonej prawidłowości.
- Wartości miary spójności informacji (rys. 6d) wykazują wyższe wartości dla wykonań korzystnych problemu strzyżenia trawnika i regresji, dla słabszych wykonań problemu mrówkowego zauważalna jest większa zmienność miary, bez określonej prawidłowości.

Przeprowadzona analiza zmienności miar entropii potwierdza wniosek z podpunktu 5.1, że miary entropii rozproszenia bloków funkcyjnych, entropii łącznej, spójności informacji oraz entropii warunkowej rozproszenia bloków względem przystosowania wykazują podobny przebieg zmienności dla wszystkich badanych problemów. Dla każdej z tych miar warto rozważyć możliwość wykorzystania w metodzie kontroli przebiegu ewolucji, która na podstawie wartości miar i zmian wartości współczynnika korelacji między nimi pozwoli wnioskować o powodzeniu procesu ewolucji. Metoda taka mogłaby być niezależna od badanego problemu.

7. Podsumowanie i propozycja dalszych badań

Przeprowadzone badania wskazują na ważność zapewnienia wysokiej wartości wybranych miar entropii dla generacji od szóstej do dwudziestej. Powyżej dwudziestej generacji wartości entropii zmniejszają się, co może mieć związek z dostosowywaniem się populacji do najlepszych osobników. Niezbędne jest zaplanowanie dalszych eksperymentów potwierdzających wyciągnięte wnioski.

Szczególnie korzystne dla badania przebiegu populacji programowania genetycznego są następujące miary: entropii rozproszenia bloków funkcyjnych, entropii łącznej, spójności informacji, entropii warunkowej rozproszenia bloków funkcyjnych względem przystosowania.

LITERATURA

1. Koza J. R.: Genetic Programming: On the Programming of Computers by Means of Natural Selection. Complex Adaptive Systems, MIT Press, Cambridge, MA, 1992.
2. Koza J. R.: Genetic Programming II: Automatic Discovery of Reusable Programs. Complex Adaptive Systems, MIT Press, Cambridge, MA, 1994.
3. Koza J. R., Bennett III F. H., Andre D., Keane M. A.: Genetic Programming III: Automatic Darwinian Invention and Problem Solving. Morgan Kaufmann, San Francisco 1999.
4. Shannon C. E: A mathematical theory of communication. Bell System Technical Journal 27 1948, s. 379÷423, 623÷653.
5. Cover T. M., Thomas J. A.: Elements of Information Theory, Wiley, New York 1991, s. 16÷24.
6. Dawkins R.: Fenotyp rozszerzony. Prószyński i S-ka SA, Warszawa 2003, s. 38÷39, 62÷63.
7. Miller, J. F.: An Empirical Study of the Efficiency of Learning Boolean Functions using a Cartesian Genetic Programming Approach, GECCO 1999.
8. Langdon W. B., Poli, R.: Foundations of Genetic Programming. Springer-Verlag, 2002.
9. McPhee N. F., Hopper N.J.: Analysis of genetic diversity through population history. In W. Banzhaf et al., editors, Proceedings of the Genetic and Evolutionary Computation Conference, Morgan Kaufmann, Florida, USA, 1999, s.1112÷1120.
10. Keijzer M.: Efficiently representing populations in genetic programming. In P. J. Angeline and K. E. Kinnear, Jr., editors, Advances in Genetic Programming 2, chapter 13, MIT Press, Cambridge, MA, USA, 1996, s. 259÷278.
11. Langdon W. B.: Data Structures and Genetic Programming: Genetic Programming + Data Structures = Automatic Programming!. Vol. 1 of Genetic Programming. Kluwer, Boston 1998.
12. Burke E., Gustafson S., Kendall G.: Genetic Programming Search Behaviour: Viewing Population Dynamics with Diversity Measures. Computer Science Technical Report No. NOTTCS-TR-2003-1, 2003.
13. Ryan C.: Pygmies and civil servants. In K.E. Kinnear, Jr., editor, Advances in Genetic Programming, MIT, 1994, s. 243÷263.
14. Ekárt A., Németh S.: A metric for genetic programs and fitness sharing. In R. Poli et al., editors, Proceedings of the European Conference on Genetic Programming, volume 1802 of LNCS, Edinburgh, 2000. Springer-Verlag, s. 259÷270.
15. Luke S., ECJ: A Java-based evolutionary computation and genetic programming system <http://www.cs.umd.edu/projects/plus/ecj/>.

16. Miller J. F., Thomson P.: Cartesian Genetic Programming. Red. R. Poli, W. Banzhaf, W. B. Langdon, J. F. Miller, P. Nordin, T. C. Fogarty, Proceedings of the Third European Conference on Genetic Programming (EuroGP2000). LNCS, Vol. 1802, Springer-Verlag, Heidelberg 2000, s. 121÷132.
17. Poli. R.: Evolution of Graph-like Programs with Parallel Distributed Genetic Programming. Red. T. Back, Genetic Algorithms: Proceedings of the Seventh International Conference, s. 346÷353, Michigan State University, East Lansing, MI, USA. Morgan Kaufmann, 1997.
18. Rosca J. P.: Entropy-driven adaptive representation. Red. J. P. Rosca, Proceedings of the Workshop on Genetic Programming: From Theory to Real-World Applications, Tahoe City, California, USA, 1995.
19. Shalizi R., Camperi M. F., Klinkner K. L.: Discovering functional communities in dynamical networks. ICML, 2006.
20. Teller A., Veloso M.: Algorithm evolution for face recognition: What makes a picture difficult. Proceedings of the International Conference on Evolutionary Computation. IEEE Press, 1995.
21. Teller A., Veloso M.: Program Evolution for Data Mining. The International Journal of Expert Systems, 1995, s. 213÷236.
22. Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., Suetens, P: Multimodality image registration by maximization of mutual information. IEEE Transactions on Medical Imaging, 1997.
23. Koronacki J., Mielniczuk J.: Statystyka dla studentów kierunków technicznych i przyrodniczych. WNT, Warszawa 2006.

Recenzent: Dr Wojciech Wieczorek

Wpłynęło do Redakcji 14 maja 2008 r.

Abstract

Genetic programming is a metaheuristic search method that uses a population of computer programs and a search strategy based on biological evolution. The goal of this thesis is to understand the dynamics of genetic programming that encourages efficient and effective search. Toward this goal, the research focuses on an important property of genetic programming search: the population diversity.

Chapter 1 presents a survey of significant diversity measures in genetic programming literature. Chapter 2 describes Cartesian Genetic Programming in which programs are repre-

sented by linear integer chromosomes in the form of connections and functionalities of a rectangular array of primitive functions. Chapter 3 introduces several new measures of population diversity which are based on entropy. Chapter 4 presents the regression problem, the ant problem and the lawnmower problem. Chapter 5 describes the experiment's environment. Chapter 6 examines measures of diversity to understand their importance and relationship with performance in genetic programming. Diversity measures are investigated in order to gain a deeper understanding of the algorithm's search behavior. Chapter 7 describes what measures are likely to be important for improving search and how successful populations create a dynamic that is suitable for the evolutionary process.

Adres

Renata ADAMCZYK: ABG S.A., ul. Wita Stwosza 7, Katowice, Polska,
renata.adamczyk@abg.com.pl .