

ALGORYTMY EWOLUCYJNE W BADANIACH ANKIETOWYCH

Zygmunt MAZUR¹, Janusz PEC²

¹Wydział Informatyki i Zarządzania, Politechnika Wrocławska, Wrocław; Zygmunt.Mazur@pwr.edu.pl

²Główny Urząd Statystyczny, Warszawa; J.Pec@stat.gov.pl

Streszczenie: W artykule przedstawiono możliwości, jakie stawiają do dyspozycji techniki ewolucyjne w obszarze badań ankieterskich. Rozszerzono je również poprzez wprowadzenie dwóch dodatkowych operatorów genetycznych. Poza znanymi z literatury przedmiotu operatorami genetycznymi, takimi jak krzyżowanie, mutacja czy inwersja, autorzy niniejszego artykułu wprowadzają dwa nowe operatory – operator „wzmocnienia genowego” oraz operator „osłabienia genowego” – umożliwiające zwiększenie stopnia wiarygodności informacji otrzymywanych przez ankietowanych. Zaproponowano również algorytm eliminacji zdublowanych chromosomów, co ma wpływ na przebieg pracy podstawowego algorytmu genetycznego poprzez zmniejszenie jego pracochłonności obliczeniowej. Zdaniem autorów, zaproponowane rozwiązania, zaakceptowane i uwzględnione w metodyce badań ankietowych, wpłyną na poprawę jakości i wiarygodności pozyskanych danych.

Słowa kluczowe: algorytm ewolucyjny, populacja, badanie ankietowe, badania statystyczne.

EVOLUTIONARY ALGORITHMIC TECHNIQUES IN THE FIELD OF QUESTIONNAIRE SURVEYS

Abstract: This article presents the possibility of using evolutionary algorithmic techniques in the field of questionnaire surveys. Also it has been extended through the introduction of two additional genetic operators outside the known from literature such as: cross-over, mutation or inversion. The newly defined operators in this article are “hardening of the gene” and “weakening of the gene” in order to increase the degree of reliability of the information received by the interviewees. The new algorithm is given in order to eliminate duplication of the chromosomes. This affects the process of modified basic genetic algorithm by reducing its workload. According to the authors the proposed solutions accepted and included in the methodology of the surveys will improve the quality and reliability of the data.

Keywords: evolutionary algorithm, population, questionnaire survey, statistical research.

1. Wprowadzanie

Jednym z problemów wykorzystania sieci ankietarskich do badań statystycznych prowadzonych przez różne instytucje komercyjne, ale również przez statystykę publiczną, jest problem wiarygodności i bezpieczeństwa zebranej informacji. Potrzeba globalnego bezpieczeństwa informacyjnego jest, jak podaje znany statystyk i ekspert w dziedzinie struktury informacyjnej państwa, były Prezes GUS prof. Józef Oleński [1], „problemem, przed którym stoją państwa i organizacje międzynarodowe” oraz również istotne jest „zapewnienie nadzoru i kontroli nad środowiskiem informacyjnym, jakie tworzą globalne infrastrukturalne systemy i procesy informacyjne. Chodzi przede wszystkim o kontrolę nad jakością i społeczną użytecznością informacji upowszechnianych w globalnej infrastrukturze informacyjnej”. W pracy [1] pisze, że „...względem globalnego środowiska informacyjnego powinno się wprowadzić zasady, normy, narzędzia kontroli jakości informacji stosując podejście podobne do tego, jakie udało się zastosować w odniesieniu do ochrony środowiska naturalnego. Upowszechnianie informacji niespełniających kryteriów jakościowych w środowisku informacyjnym i globalnych systemów infrastrukturalnych powinno być zabronione w ramach konwencji ogólnoswiatowej przyjętej przez społeczność międzynarodową, jako konkretyzacja Powszechnej Deklaracji Praw Człowieka przyjętej przez ONZ w 1948 roku w odniesieniu do zapisów dotyczących wolności i bezpieczeństwa ludzi w warunkach globalnego społeczeństwa informacyjnego”.

Rozwiązanie problemu wiarygodności danych określanych czasem w potocznym języku, jako eliminacja „sufitologii”, stanowi podstawę rzetelności otrzymanych wyników – publikowanych później w różnych mediach i profesjonalnych opracowaniach, artykułach – w tym badawczych. W celu poprawy „wiarygodności danych” autorzy niniejszego opracowania zaproponowali wykorzystanie technologii algorytmów ewolucyjnych, odpowiednio zmodyfikowanej przez autorów, przystosowanej do badań ankietowych.

W klasycznej teorii algorytmów genetycznych mamy do czynienia z trzema podstawowymi operacjami na ciągach kodowych zwanych chromosomami (uporządkowane ciągi genów) używanymi najczęściej w postaci zapisu binarnego jako ciągi zerojedynkowe. Są to: operacja krzyżowania (ang. *crossover*), operacja mutacji (ang. *mutation*) oraz rzadziej stosowana operacja inwersji (ang. *inversion*). W celu przystosowania algorytmów genetycznych do badań statystycznych autorzy niniejszego opracowania wprowadzają dwa dodatkowe operatory genetyczne zwane „wzmocnieniem genowym” i „osłabieniem genowym”. Wszystkie wymienione tutaj operatory genetyczne zostaną omówione w rozdziale poświęconym ich charakterystyce. Wprowadzenie dwóch dodatkowych operatorów pozwala uwzględnić problemy związane z bezpieczeństwem informacji pojmowanym tutaj w sensie

ontologicznym (semantycznym)¹, w tym z problemem wiarygodności danych zawartych w chromosomach. W opisie technik algorytmów ewolucyjnych będziemy używać powszechnie stosowanej w literaturze przedmiotu terminologii znanej w genetyce, a więc wyrażeń takich jak: populacja, osobniki (zamiennie używa się też terminu „organizmy”), chromosomy, gen, genotyp, fenotyp, allel, locus.

Podajemy niżej definicje używanych podstawowych pojęć:

- Populacja – zbiór osobników o danej liczebności.
- Osobnik – członek populacji zakodowany w postaci chromosomów.
- Gen – pojedynczy element chromosomu – w klasycznym algorytmie genetycznym to 0 lub 1.
- Chromosom – uporządkowany łańcuch genów.
- Genotyp – zespół wszystkich genów, który warunkuje dziedziczne właściwości danego organizmu, czyli struktura składająca się z chromosomów.
- Fenotyp – zespół dostrzegalnych cech danego organizmu (anatomicznych, morfologicznych, biochemicznych i fizjologicznych), które zostały wykształcone pod wpływem czynników dziedzicznych (genotyp), modyfikowanych przez warunki środowiskowe.; w naszym przypadku – zbiór parametrów zadania będący jego rozwiązaniem w przestrzeni poszukiwań (ang. *search points*).
- Allel – alternatywna forma genu, jedna z dwóch (bądź więcej) postaci danego genu reprezentująca wartość cechy lub jej wariant.
- Locus – położenie danego genu K w łańcuchu, czyli w chromosomie, oznaczone l_K .

W następnym rozdziale wyjaśnimy na przykładach działanie operatorów genetycznych. W artykule zakładamy kodowanie binarne chromosomów – jak w klasycznym algorytmie genetycznym.

2. Charakterystyka operatorów genetycznych

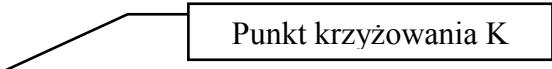
Na początku omówimy podstawowe operatory genetyczne.

Operator krzyżowania O_K

Założmy, że mamy dwa następujące chromosomy zapisane w postaci zer i jedynek. Jeżeli w chromosomie mamy L genów, to wybór punktu K krzyżowania z danym prawdopodobieństwem p_K musi spełniać oczywisty warunek $l_K \in [1, L-1]$.

¹ Pojęcie bezpieczeństwa semantycznego jest również używane w kryptologii, ale jest tam inaczej rozumiane – zgodnie z jej formalną definicją w tym obszarze wiedzy.

Założmy dalej, że oba chromosomy mają po 14 genów, w związku z czym losujemy z pewnym prawdopodobieństwem p_k (z reguły zakładamy $0,5 \leq p_k \leq 1$) liczbę z przedziału $[1,13]$ w celu określenia położenia punktu krzyżowania i niech to będzie liczba 6.



Punkt krzyżowania K

Chromosom $ch_A = [001101 | 11100010]$

Chromosom $ch_B = [011101 | 10100111]$

Po dokonaniu operacji krzyżowania otrzymujemy nową parę potomków.

$ch_{A'} = [00110110100111]$

$ch_{B'} = [01110111100010]$

Należy również wspomnieć, że istnieją inne rodzaje krzyżowań znane w literaturze, takie jak krzyżowanie wielopunktowe (ang. *multiple-point crossover*) czy jednostajne (ang. *uniform-crossover*).

Operator mutacji O_m

Rozważmy chromosom $ch_A = [00110111100010]$ z poprzedniego przykładu. Przy tej operacji zakłada się, że prawdopodobieństwo mutacji p_m jest bardzo małe i z reguły spełniona ono następujący warunek: $0 \leq p_m \leq 0,1$. Przyjmijmy wartość $p_m = 0,09$. Losujemy 14 liczb z przedziału $[0,1]$ z dokładnością do dwóch miejsc po przecinku i założmy, że otrzymaliśmy następujące wyniki: 0,11 0,31 0,07 0,27 0,13 0,33 0,29 0,71 0,61 0,19 0,55 0,05 0,41 0,15. Mutacja zmienia bit na przeciwny, czyli 0 na 1 i 1 na 0. Zamieniamy więc bity na przeciwne na pozycjach 3 i 12 w chromosomie ch_A , bowiem tylko na tych miejscach spełniony jest warunek dotyczący prawdopodobieństwa mutacji p_m . W rezultacie otrzymujemy nowy zmutowany chromosom $ch_{A'}$, tj. o zmienionych allelach na pozycjach 3 i 12:

$$ch_A = [00110111100010] \rightarrow [00010111100110] = ch_{A'}$$

Operator inwersji O_{in}

Inwersja działa na pojedynczym chromosomie i polega na zmianie kolejności alleli między dwoma losowo wybranymi pozycjami (locus) chromosomu. Operacje inwersji przeprowadzimy na tym samym chromosomie $ch_A = [00110111100010]$. Założmy, że wylosowano następujące pozycje: 5, 11. Otrzymujemy chromosom $ch_{A'}$, jak niżej

$$ch_A = [00110111100010] \rightarrow [00110001111010] = ch_{A'}$$

Wprowadzimy teraz dwa nowe operatory genetyczne ułatwiające ocenę informacji zawartej w ankietach. Będą to operator „wzmocnienia genowego” i „osłabienia genowego”. Te dwa operatory zostały wprowadzone przez autorów niniejszego artykułu celowo na potrzeby ankietowych badań statystycznych.

Operator „wzmocnienia genowego” działa na określonym podzbiore populacji. Mając daną liczbę osobników populacji losujemy pewien interesujący nas podzbiór osobników

z zadaniem przez nas zmiennym prawdopodobieństwem, za pomocą generatora liczb losowych. W badaniach statystycznych, w tym ankietarskich, możemy poprzez dobór odpowiednich wartości prawdopodobieństwa wyboru danego osobnika lub podzbioru osobników manipulować charakterystyką próbki, stosownie do naszych potrzeb. Operator „wzmocnienia genowego” umożliwi nam identyfikację potomka lub grupę potomków osobników w populacji o podobnej strukturze genowej, zwiększając tym samym tym samym stopień pewności zebranej jednostkowej informacji o obiekcie w rozumieniu technologii obiektowej, a więc traktowanego jako wyodrębniony element świata materialnego lub abstrakcyjnego. Im więcej jest w populacji wyjściowej (rodziców) chromosomów podobnych – o pokrywających się na tym samym l_k (locus) wartościach bitów – tym bardziej wiarygodna staje się informacja o danym obiekcie (zwiększa się też liczba bardziej wiarygodnych potencjalnych rozwiązań) zebrana jedną z metod opisanych w 3. rozdziale artykułu (przy okazji omawiania modelu sieci ankietarskiej). Gwarantuje to w pewnej mierze niezależność źródeł informacji (różne agencje ankietarskie stosujące różne techniki badań). Uwzględniając długość chromosomu L (ilość bitów) możemy oceniać podobieństwo genowe dwóch chromosomów jako wskaźnik $wsk_p = \frac{s}{L} \cdot 100\%$, gdzie liczba s to ilość pokrywających się bitów w obu chromosomach. Można oczywiście dla danego zbioru informacji o obiekcie, czyli zbioru chromosomów powiedzmy w ilości n , obliczyć dla każdej z $\binom{n}{s}$ par wartość maksymalną lub minimalną, lub średnią (np. arytmetyczną, geometryczną) albo inne parametry statystyczne charakteryzujące wartości średnie. Wydaje się, że wprowadzenie do rozważań operacji „wzmocnienia genowego” ma intuicyjny sens.

Operację tę można wykonać na wielu chromosomach, co oczywiście z jednej strony intuicyjnie sugeruje zwiększenie stopnia wiarygodności danej informacji o obiekcie zawartej w otrzymanym potomku, ale z drugiej strony poszerza przestrzeń poszukiwań najlepszego chromosomu poprzez zwiększenie ich liczby.

Niżej podamy opis działania operatora „wzmocnienia genowego”.

Operator „wzmocnienia genowego” O_{wg}

Rozważmy nasze przykładowe dwa chromosomy:

$$ch_A = [00110111100010] \text{ i } ch_B = [01110110100111].$$

Operator „wzmocnienia” generuje chromosomy potomne o następującym schemacie $Sch_{p_{wg}} = [0X11011X100X1X]$, gdzie X oznacza 0 albo 1, a prawdopodobieństwo $p_w \leq \frac{1}{2} \mp \varepsilon$, ($0 \leq \varepsilon \leq 0,01$). Jak widać, chromosomy potomne powielają wartości bitów na pozycjach, na których wartości „chromosomów rodziców” się pokrywają (są identyczne), na pozostałych miejscach wartości bitów (genów) chromosomu potomnego występują dowolne wartości ze zbioru dwuelementowego $\{0,1\}$. Jeżeli rozważymy dwa chromosomy o długości L , o m powielających się bitach, to w wyniku zastosowania operatora

„wzmocnienia genowego” uzyskujemy schemat potomka o $L - m$ pozycjach z wartościami X . Tak więc, w takim przypadku uzyskujemy 2^{L-m} chromosomów potomnych. W naszym przykładzie uzyskujemy 2^{14-10} , ($2^4=16$) potomków, czyli genotyp $Sch_{p_{wg}}$, a mianowicie:

- | | |
|---------------------|----------------------|
| 1. [00110110100010] | 9. [01110110100010] |
| 2. [00110110100011] | 10. [01110110100011] |
| 3. [00110110100110] | 11. [01110110100110] |
| 4. [00110110100111] | 12. [01110110100111] |
| 5. [00110111100010] | 13. [01110111100010] |
| 6. [00110111100011] | 14. [01110111100011] |
| 7. [00110111100110] | 15. [01110111100110] |
| 8. [00110111100111] | 16. [01110111100111] |

Nowy zdefiniowany operator „genowego osłabienia” ma również na celu poprawę wiarygodności zbioru chromosomów w populacji potomnej. Załóżmy jak poprzednio przy omawianiu „wzmocnienia genowego”, że otrzymujemy plik informacji o obiekcie zapisanej w formie zbioru genów otrzymanych z różnych źródeł (np. od niezależnych agencji ankierskich). Jeżeli w trakcie analizy tego zbioru chromosomów stwierdzimy ich dużą rozbieżność, przejawiającą się jako niski stopień pokrywania się bitów między każdą parą chromosomów (mała wartość wskaźnika podobieństwa wsk_p), to rzeczą słuszną wydaje się użycie operatora „osłabienia genowego” w celu uzyskania bardziej wiarygodnych informacji o obiekcie od jego gestora. Otrzymywanie bowiem zróżnicowanych informacji, czyli zbioru rozbieżnych chromosomów od gestora obiektu dla poszczególnych ankierów, sugeruje celowe przekłamania. Rozsądne wydaje się w tym przypadku przyjęcie założenia, że pokrywanie się bitów będzie wskazywało na potrzebę zmiany ich wartości 0 na 1 lub odwrotnie, jeżeli w populacji rodzicielskiej nie został spełniony warunek zatrzymania (patrz schemat na Rysunku 2.) dla żadnego chromosomu z tej populacji. W populacji potomnej uzyskujemy zatem większą pewność, iż posiadamy w niej optymalne lub suboptymalne rozwiązanie.

Pozwoli to w jakimś stopniu zwiększyć wiarygodność nowo otrzymanego zbioru chromosomów. Taki jest opisany tutaj bezpośredni, a zarazem intuicyjny, sens tej operacji genowej. Jak wspomniano wyżej, przy operacji „osłabienia genowego” nie dublujemy rekordów, mimo że znacznie zwiększamy przestrzeń przeszukiwania najlepszego chromosomu spełniającego warunek zatrzymania. W tym przypadku zmniejszamy oczywiście złożoność obliczeniową zmodyfikowanego modułu pokazanego na Rysunku 2.

Operator „osłabienia genowego” O_{og}

Działanie nowego operatora „osłabienia genowego” jest w pewnym sensie odwrotne do działania operatora „wzmocnienia genowego”. Dla chromosomu rodziców ch_A i ch_B o długości L powstaje 2^{L-m} chromosomów potomnych. Schemat $Sch_{p_{og}}$ posiada na pozycjach

l_k , na których powielają się wartości bitów (genów), tylko jedną z wartości $\in \{0,1\}$ według reguły:

1 – jeżeli wartość bitów rodziców wynosiła 0

0 – jeżeli wartość bitów rodziców wynosiła.

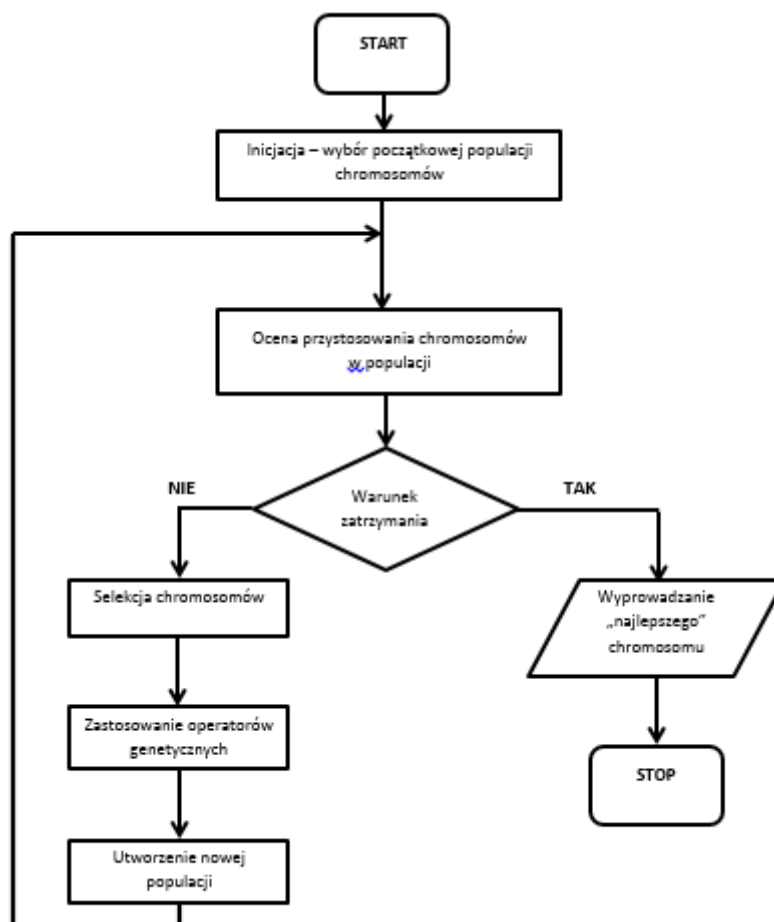
Natomiast na pozostałych l_k (element X w schemacie $Sch_{p_{og}}$) pozostawia bez zmian wartości bitów ze zbioru $\{0,1\}$. Rozpatrzmy dwa chromosomy rodziców $ch_A = [00110111100010]$ oraz $ch_B = [01110110100111]$ z ostatniego przykładu. W wyniku działania operatora „osłabienia genowego” otrzymamy następujący schemat $Sch_{p_{og}} = [1X00100X011X0X]$. Otrzymujemy, zatem ich liczbę równą $2^4=16$ chromosomów potomnych.

Oto one:

- | | |
|---------------------|----------------------|
| 1. [11001000011101] | 9. [10001001011101] |
| 2. [11001000011100] | 10. [10001001011100] |
| 3. [11001000011001] | 11. [10001001011001] |
| 4. [11001000011000] | 12. [10001001011000] |
| 5. [11001001011101] | 13. [10001000011001] |
| 6. [11001001011100] | 14. [10001000011000] |
| 7. [11001001011001] | 15. [10001000011101] |
| 8. [11001001011000] | 16. [10001000011100] |

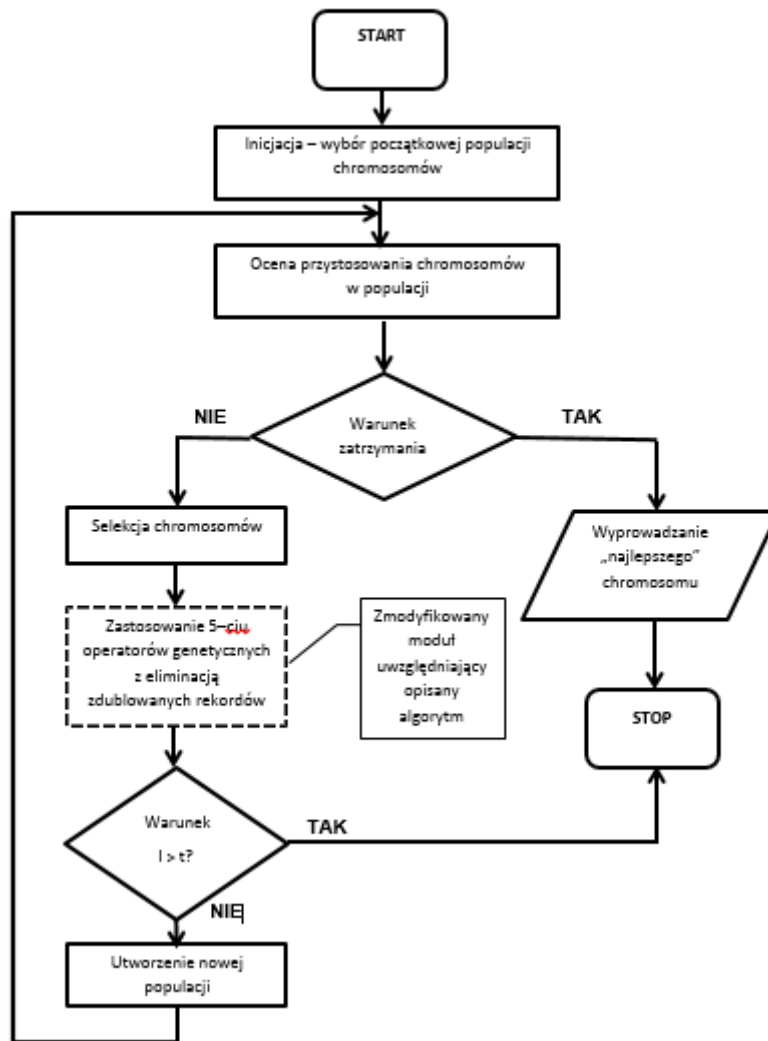
Łatwo zauważyć, że z definicji operatorów „wzmocnienia” i „osłabienia genowego” wynika fakt, iż żaden z dwóch powyższych schematów nie może wyprodukować dwóch takich samych chromosomów. Natomiast z dwóch potomków uzyskanych w wyniku zastosowania operatora „wzmocnienia genowego” i powtórnego lub wielokrotnego użycia tego samego operatora uzyskujemy efekt dublowania się chromosomów. Inaczej jest w przypadku „osłabienia genowego”. Wprost z definicji wynika, iż powtarzalne użycie tego operatora generuje nowe niepowtarzające się chromosomy w danej populacji. Ale oczywiście dwa inne chromosomy rodzicielskie podlegające działaniu poznanych innych operatorów genetycznych – już tak. Biorąc np. pod uwagę dwa chromosomy rodzicielskie z przytoczonego wcześniej przykładu krzyżowania i ich drugiego potomka $ch_{B'} = [01110111100010]$, po operacji mutacji na pozycjach 12 i 14 otrzymamy chromosom $ch_{B''} = [01110111100111]$, który jest 16. chromosomem na liście potomków „wzmocnienia genowego” tych chromosomów rodzicielskich w pokazanym wyżej przykładzie. Dążąc do ograniczenia przestrzeni poszukiwania najlepszego chromosomu, ze względu na potrzebę zmniejszenia złożoności obliczeniowej rozpatrywanego zagadnienia, należy eliminować dublujące się chromosomy. Dalej w artykule podano prosty algorytm wykonywania tej operacji w zbiorze chromosomów poddanych operacjom genetycznym, przy założeniu, że wybraną metodą selekcji chromosomów w podanym podstawowym schemacie blokowym

algorytmu genetycznego [2] (Rysunek 1) jest jedną z bardziej popularnych i omawianych w literaturze metod – „metoda koła ruletki” (ang. *roulette-wheel selection*).



Rysunek 1. Podstawowy schemat blokowy algorytmu genetycznego.

Istnieją oczywiście inne znane teorii i praktyce metody selekcji, takie jak selekcja rankingowa, turniejowa czy metoda „stłoczenia”, zapewniająca zachowanie różnorodności populacji. Tak więc, po uwzględnieniu operacji eliminowania zdublowanych chromosomów otrzymujemy zmodyfikowany podstawowy algorytm genetyczny, jak pokazano na Rysunku 2. Na koniec dygresji o proponowanym algorytmie warto wspomnieć o warunku zatrzymania (krok 10. algorytmu), którego postać może być wyrażona albo jako wartość skalarna, albo jako wektor. Ponieważ, jak zaznaczyliśmy na początku artykułu, rozważamy tutaj głównie kodowanie binarne, należy również wspomnieć o kodowaniu niebinarnym chromosomów, gdzie operacje genetyczne są inaczej definiowane w oparciu o struktury drzewiaste.



Rysunek 2. Zmodyfikowany podstawowy schemat blokowy algorytmu genetycznego.

W przypadku kodowania niebinarnego, z technikami ewolucyjnymi ściśle jest związane programowanie genetyczne spopularyzowane przez Johna Kożę na początku lat 90. XX wieku. Programy przedstawiane są za pomocą drzew składniowych (ang. *parsing tree*). Wykorzystując jeden z najbardziej popularnych języków programowania dla kodowania algorytmów genetycznych, jakim jest LISP, możemy właśnie w tym języku stosunkowo prosto zapisać wspomniane już „struktury drzewiaste”. Populacja osobników może osiągać nawet setki milionów, więc problem redukcji złożoności obliczeniowej jest tutaj istotny. Wiele problemów można próbować rozwiązać za pomocą programowania równoległego, bowiem w realnych zastosowaniach przestrzeń potencjalnych rozwiązań jest zwykle tak duża, że jest praktycznie niemożliwe sprawdzenie w dostatecznie krótkim czasie wszystkich możliwych rozwiązań i wybranie rozwiązania optymalnego lub jemu bliskiego z określonym błędem tolerancji (tzw. suboptymalnego). Gama zastosowań programowania ewolucyjnego jest dość szeroka. Począwszy od harmonogramowania i planowania procesów przemysłowych, projektowania topologii sieci elektrycznych, poprzez przemysł chemiczny,

a skończywszy na wykorzystaniu tej techniki w medycynie – diagnozowanie i prognozowanie, przetwarzanie obrazów uzyskanych tomografią, rezonansem magnetycznym czy PET-em, wspomniane już wyżej planowanie i harmonogramowanie. Liczba zastosowań w przyszłości na pewno będzie wzrastać.

Jeżeli struktura ankiety zawiera odpowiedzi na pytania – jak to zostanie zdefiniowane w opisie modelu sieci ankietarskiej w następnym rozdziale (tj. 5) – i jeżeli przeznaczymy odpowiednią liczbę miejsc na zapis informacji w chromosomie o długości L (tzn. stworzymy pola o pewnej długości), to dla każdej z nich mamy relację – odwzorowanie $\phi: X \rightsquigarrow Y$, gdzie X stanowi zbiór możliwych odpowiedzi, a Y zbiór ankietowanych, natomiast strzałka „ \rightsquigarrow ” oznacza multifunkcję. Generalnie jest to multifunkcja, czyli odwzorowanie wielowartościowe, jako uogólnienie pojęcia funkcji poprzez dopuszczenie przyporządkowania każdemu elementowi dziedziny więcej niż jednego elementu przeciwdziedziny, bowiem ankietowany podmiot może udzielić różnych odpowiedzi na te same pytania, w zależności od tego, kto je zadaje. Mogą tu wchodzić w rachubę różne czynniki, w tym psychologiczne.

3. Model sieci ankietarskiej

Rozważmy następujący model sieci ankietarskiej. Załóżmy, że spełnione są poniższe warunki:

1. Dane jest:

- K podmiotów ankietowanych; ankiety dotyczące M różnych spraw,
- N niezależnych agencji ankietarskich/ankietowanych – poszerzenie bazy danych etapu inicjacji w schemacie zmodyfikowanego podstawowego algorytmu genetycznego (Rysunek 2.).

2. Ankietujący uwzględniają specyfikę działalności ankietowanego podmiotu, co powoduje zwiększenie stopnia wiarygodności otrzymanych informacji. Struktura ankiety składa się z pól o określonej liczbie komórek, z których każda zawiera bit „0”, bit „1” lub wartość null i posiada następujące kategorie odpowiedzi:

- Odpowiedzi: „TAK”, „NIE”, „NIE DOTYCZY”. Ostatnia odpowiedź jest rejestrowana w chromosomie jako pole z wartością typu null. Jest to jednokomórkowe pole, które nie jest uwzględniane przy operacjach genowych, a przy obliczaniu fenotypu nie jest brane pod uwagę. Zwiększa ono tylko formalnie długość L chromosomu i jest pomijane we wszystkich modułach zmodyfikowanego podstawowego algorytmu genetycznego podczas jego przebiegu zgodnie ze schematem przedstawionym na Rysunku 2.
- Odpowiedzi typu liczba.
- Odpowiedzi będące listą obiektów – listą nazw. Ten rodzaj odpowiedzi można łatwo sprowadzić do odpowiedzi typu „TAK”, „NIE”, poprzez podanie listy przez ankietowanego

i zaznaczenie przez ankietowanego, czy jest w jakiegokolwiek relacji z danym obiektem (nazwą).

3. Prawdopodobieństwo udzielenia informacji zgodnej z prawdą przez podmiot k , $1 \leq k \leq K$ ankietowani n dotyczącej tematu t , $1 \leq t \leq M$, $1 \leq n \leq N$ wynosi $p_{k,t,n}$.

Możemy potraktować to jako dyskretny rozkład prawdopodobieństwa na zbiorze chromosomów aktualnej rozpatrywanej populacji. W przypadku wielkiej liczby chromosomów w zbiorze Z chromosomów możemy rozważać rozkłady ciągłe. Warto tutaj zauważyć, że znając moc $|Z|$ zbioru Z , który jest określony w kroku 7. podanego niżej algorytmu, możemy dzięki wprowadzonym przez nas operacjom „wzmocnienia” i „osłabienia genowego” oszacować z góry punktowo prawdopodobieństwa $p_{k,t,n}$.

Umożliwia nam to oszacowanie górnej granicy maksymalnej liczby chromosomów uzyskanych za pomocą operatorów „wzmocnienia” lub „osłabienia”. Oszacowanie to zostało podane w ostatnim rozdziale niniejszego artykułu. Znając punktowe oszacowania prawdopodobieństw $p_{k,t,n}$ możemy starać się o interpolację funkcji, która będzie szacować z góry szukaną funkcję gęstości prawdopodobieństwa jedną z wielu znanych metod stosowanych w analizie numerycznej.

4. Koszt pojedynczego badania ankietowego dla danej k -tej agencji ankietarskiej podmiotu i dla badania ankietowego dotyczącego tematu m wynosi $c_{k,m,n}$.

Można go uważać za interesujący nas np. pewien parametr θ . Biorąc pod uwagę pewien rozkład prawdopodobieństwa, o którym była mowa wyżej, możemy być zainteresowani wyznaczeniem estymatora $\hat{\theta}$ parametru θ metodą największej wiarygodności. Jest dość oczywisty fakt, że funkcja gęstości prawdopodobieństwa opisująca wiarygodność otrzymywanych danych jest w naszym przypadku w jakimś stopniu zależna od kosztów badań.

Zakładamy tutaj, że koszty ankiet dotyczących różnych spraw (tematów) są różne, co jest zrozumiałe. Jest też zrozumiałe, że dane badanie – ankieta – może spowodować ponoszenie różnych nakładów (kosztów) przez poszczególne agencje ankietarskie, choćby ze względu na stosowanie różnych technik, o których mowa w następnym punkcie.

5. Odpowiedzi mogą być udzielane metodami ilościowymi, takimi jak:

- metoda PAPI (z ang. *Paper and Pencil Interview*) – termin określający zbiorczo techniki badań ilościowych, bazujące na konwencjonalnej, papierowej wersji kwestionariuszy badawczych, w których pisemnie zaznaczane są odpowiedzi respondentów;
- metoda CAPI (od ang. *Computer-Assisted Personal Interviewing* – wywiad osobisty ze wspomaganiami komputerowymi) – technika badań ilościowych, w której bezpośredni kontaktowi ankietera z respondentem towarzyszy komputer (zwykle przenośny), zaopatrzony w specjalistyczne oprogramowanie wspomagające realizację wywiadu;

- metoda CASI (z ang. *Computer-Assisted Self Interviewing* – komputerowa ankieta wypełniana samodzielnie przez respondenta) – technika badań ilościowych zbliżona do systemu CAPI;
- metoda CATI (z ang. *Computer-Assisted Telephone Interviewing* – wywiad telefoniczny ze wspomaganie komputerowym) – technika badań ilościowych, w której pozyskiwanie informacji od respondentów odbywa się za pośrednictwem łączy telefonicznych i przy wykorzystaniu specjalistycznego oprogramowania komputerowego;
- metoda CAWI (z ang. *Computer-Assisted Web Interviewing* – ankieta internetowa nadzorowana przez system komputerowy) – popularyzująca się technika badań ilościowych, w której pytania kwestionariuszowe pobierane są ze strony internetowej organizatora badania i przekazywane za pośrednictwem sieci do dowolnego punktu, w którym znajduje się respondent wraz z komputerem podłączonym do Internetu

lub metodami uważanymi za jakościowe:

- metoda FGI (z ang. *Focus Group Interviews*) – jedna z najpopularniejszych metod badań jakościowych, polegająca na wspólnej dyskusji grupy respondentów / uczestników wywiadu na zadany z góry temat lub grupę tematów;
- metoda IDI (z ang. *Individual In-Depth Interviews*) – jedna z bardziej popularnych metod badań jakościowych, polegająca na szczegółowej, wnikliwej rozmowie z informatorem / respondentem, której celem jest dotarcie do jakichś precyzyjnych informacji, poszerzenie wiedzy związanej z tematem badania;
- metoda obserwacji – polegająca na śledzeniu przez badacza określonych zachowań respondentów w warunkach naturalnych (galerie handlowe, imprezy masowe);
- precyzyjnie przygotowany wywiad na podstawie przemyślanego planu, często we współpracy z projektodawcą badania – przygotowany scenariusz dla indywidualnego respondenta lub grupy respondentów.

Oczywiście możemy łączyć te metody stosując w ten sposób techniki mieszane w badaniach ankietarskich, co być może uwiarygadnia wyniki, ale oczywiście z reguły podraża koszty badań. Dokładniejsze opisy tych badań można znaleźć choćby na internetowych stronach urzędów statystycznych, zwłaszcza na stronie GUS www.stat.gov.pl w zakładce poświęconej informacji o badaniach ankietowych oraz na stronie urzędu statycznego w Łodzi, która posiada wyspecjalizowaną komórkę – Ośrodek Warunków Życia i Badań Ankietowych. Statystyka publiczna poprzez swoje służby prowadzi społeczne i rolnicze badania ankietowe, które ze względu na częstotliwość realizacji badań dzielimy na badania stałe, cykliczne i jednorazowe (*ad hoc*).

6. Podział chromosomu o długości L , przypisanego danej ankiecie, na pola zerojedynkowe odpowiadające pytaniom i niech będzie ich s , pole z wartością „null” oraz

p pól F_1, F_2, \dots, F_p zawierające liczby podawane przez podmiot ankietowany o długościach odpowiednio $|F_i|$, $i = 1, 2, \dots, S$, przy czym musi zachodzić następujący warunek:

$$s + 1 + \sum_{i=1}^{i=p} |F_p| = L.$$

Użyto w oznaczeniach symbolu mocy zbioru „| |”, gdyż długość pola jest tutaj równa liczbie elementów pola chromosomu.

Typowy przykład ankiety o przywołanej niniejszym strukturze odpowiedzi dotyczący ciągłości działania i zarządzania bezpieczeństwem informacji w instytucji rządowej będącej centralnym organem administracji państwowej jaką jest GUS, można znaleźć w pracy [3]. Podany tam przykład stara się uwzględnić specyfikę działalności centralnego organu administracji państwowej w zakresie badań statystycznych.

4. Modyfikacja algorytmu genetycznego

Przedstawimy modyfikację algorytmu genetycznego w aspekcie zaprezentowanego modelu ankietarskiego w poprzednim rozdziale.

Algorytm zmodyfikowanego modułu *Zastosowanie operatorów genetycznych*

Dane wejściowe algorytmu

5 operacji (krzyżowanie, mutacja, inwersja, „wzmocnienie genowe”, „osłabienie „genowe”)

t – liczba generacji pokoleń,

L – liczba genów w chromosomie,

LST_i – lista chromosomów otrzymana z modułu „selekcja chromosomów” (patrz Rysunek 2.).

Oznaczenia

LST_i – aktualna lista chromosomów w i -tej generacji pokolenia, $i = 1, 2, \dots, t$

p_k^{gr} – graniczne prawdopodobieństwo krzyżowania

p_m^{gr} – graniczne prawdopodobieństwo mutacji

p_{in}^{gr} – graniczne prawdopodobieństwo inwersji

p_{wg}^{gr} – graniczne podobieństwo „wzmocnienia genowego”

p_{og}^{gr} – graniczne prawdopodobieństwo „osłabienia genowego”

x – jeden z symboli ze zbioru {„k”, „wg”, „og”}

$p_i^x(ch_k, ch_j)$ – prawdopodobieństwo użycia operatora genowego dwuargumentowego O_k, O_{wg}

lub O_{og}

w i -tej generacji

$p_i^m(ch_k)$ – prawdopodobieństwo mutacji chromosomu ch_k w i -tej generacji

$p_i^{in}(ch_k)$ – prawdopodobieństwo inwersji chromosomu ch_k w i -tej generacji

Dane wyjściowe algorytmu

Lista LST_i bez zdublowanych rekordów

Krok 1.

Odebranie populacji LST_i po selekcji chromosomów oraz wartości i , jak na załączonym schemacie na Rysunku 2.

Krok 2.

Generujemy za pomocą wybranego generatora liczb pseudolosowych 5 prawdopodobieństw granicznych: $p_k^{gr}, p_m^{gr}, p_{in}^{gr}, p_{wg}^{gr}, p_{og}^{gr}$ z przedziału $[a_x, b_x] \subset [0,1]$ dla danej generacji i ; index x przy parametrach a i b oznacza dolną i górną wartość przedziału liczbowego dla danego typu operatora genetycznego, czyli x przebiega zbiór symboli $\{k, m, in, wg, og\}$ oznaczający typ operatora genetycznego O_x tj. $x \in \{k, m, in, wg, og\}$.

Krok 3.

Wyznaczamy $\binom{|LST_i|}{2}$ par chromosomów i dla każdej pary wyznaczamy, za pomocą generatora liczb pseudolosowych, zbiór trójek prawdopodobieństw $\{p_i^k, p_i^{wg}, p_i^{og}\}$. dla trzech dwuargumentowych operatorów genetycznych: O_k, O_{wg}, O_{og} , i – kolejny numer generacji. Symbol $|LST_i|$ oznacza moc zbioru LST_i .

Krok 4.

Tworzymy zbiory par chromosomów:

$Z_i^x := \{ (ch_k, ch_j) : 1 \leq k, j \leq |LST_i|, p_i^x(ch_k, ch_j) \in [a_x, b_x] \subset [0,1] \}$ dla każdego operatora dwuargumentowego: O_k, O_{wg}, O_{og} , gdzie i jest kolejnym numerem generacji pokoleń, a x przebiega zbiór symboli $\{k, wg, og\}$, tj. $x \in \{k, wg, og\}$.

Krok 5.

Dla operatorów jednoargumentowych O_m, O_{in} tworzymy w danej generacji i dwa zbiory chromosomów:

$$Z_i^m = \{ ch_k : 1 \leq k \leq |LST_i|, p_i^m(ch_k) \in [a_m, b_m] \subset [0,1] \}$$

$$Z_i^{in} = \{ ch_k : 1 \leq k \leq |LST_i|, p_i^{in}(ch_k) \in [a_{in}, b_{in}] \subset [0,1] \}$$

Krok 6.

Wykonujemy dla danej generacji i operacje genetyczne dwuargumentowe O_k, O_{wg}, O_{og} , zgodnie z ich podanymi definicjami, na parach chromosomów należących do zbiorów Z_i^x określonych w kroku 4. i dla operatorów jednoargumentowych O_m, O_{in} na chromosomach określonych w kroku 5., należących do zbiorów Z_i^m i Z_i^{in} .

Krok 7.

Dla danej generacji i tworzymy zbiór $Z = Z_i^k \cup Z_i^{wg} \cup Z_i^{og} \cup Z_i^m \cup Z_i^{in}$.

Krok 8.

Wykonujemy pętlę „for”:

for k := 1 step=1 to |Z| do:

porównujemy genotyp ch(k) z pozostałymi chromosomami zbioru Z – porównanie kolejnych łańcuchów zerojedynkowych z ch(k). Jeżeli łańcuchy są różne, to pozostawiamy oba chromosomy w zbiorze Z, w przeciwnym przypadku usuwamy jeden z nich.

enddo; .

Krok 9.

i:= i + 1

Jeżeli i > t, to przechodzimy do kroku 10.

Porządkujemy zbiór Z np. rosnąco numerując kolejne chromosomy i przepisujemy do listy, czyli w i-tej generacji tworzymy nową listę $LST_i = Z$ i przechodzimy do kroku 3.

Krok 10.

STOP zmodyfikowanego algorytmu genetycznego z uwagi na ograniczenie liczby iteracji (ze względów czasowych – przekroczenie czasu) lub przekroczenie kosztów generowania kolejnych populacji potomków.

Złożoność obliczeniowa tego algorytmu, jak łatwo zauważyć po analizie kolejnych kroków algorytmu, zależy w znacznej mierze od złożoności obliczeniowej algorytmu generatora liczb losowych.

Mając do dyspozycji określony model sieci ankietarskiej oraz zdefiniowany i zmodyfikowany podstawowy schemat blokowy algorytmu genetycznego, możemy precyzyjniej określić warunek zatrzymania. Warunek zatrzymania jest funkcją wartości przystosowania chromosomu – im większa wartość tej funkcji tym lepsza „jakość” chromosomu. Warunkiem zatrzymania może być np. limit wydatków przeznaczonych na badania ankietowe. Oczywiście można sformułować w inny sposób warunek zatrzymania. Może to być również przekroczenie limitu czasowego lub osiągnięcie maksimum funkcji przystosowania z wartościami nieujemnymi albo brak poprawy rezultatów w wyniku dalszego działania algorytmu. Jeżeli przyjmiemy, jako warunek zatrzymania, znalezienie wartości estymatora $\hat{\theta}$ jakiegoś parametru θ (np. wspomnianego już kosztu badania) skonstruowanej w pewien sposób funkcji wiarygodności, dla którego przyjmuje ona największą wartość (maximum), to oczywiście powinna ona spełniać podstawowe założenia znane z analizy matematycznej (różniczkowalność, warunki przyjmowania największej wartości w pewnym punkcie zbioru).

W przypadku, kiedy funkcja przystosowania $f(x)$ ma inną postać niż omówiona wyżej, to należy ją tak przekształcić, aby spełniała pierwotne założenia – na przykład, kiedy rozpatrujemy w programowaniu liniowym zagadnienie minimalizacji funkcji $\min f(x)$, to omawiane zagadnienie możemy łatwo sprowadzić do zagadnienia maksymalizacji funkcji: $-f(x)$.

Warunkiem zatrzymania może być też warunek najlepszego dopasowania chromosomów między dwiema kolejnymi epokami uzyskanego metodą najmniejszych kwadratów, co sugerowałoby pewną stabilizację wyników, ale wydaje się, że metoda największej wiarygodności ma tutaj przewagę ze względu na fakt, że wymaga mniej restrykcyjnych założeń.

Jest rzeczą oczywistą, że określenie warunku zatrzymania należy dopasować do specyfiki problemu, który jest przedmiotem naszej analizy. Zjawiska występujące w naturze często mają rozkład normalny $N(\mu, \sigma^2)$, gdzie μ jest wartością oczekiwaną a σ^2 wariancją o znanej funkcji gęstości $f_{\mu, \sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$. Rozkład normalny często jest wykorzystywany przy tworzeniu strategii ewolucyjnej – zwłaszcza standardowy rozkład normalny $N(0,1)$. Wykorzystuje się go ze względu na łatwość wykonywanych operacji matematycznych w metodach statystycznych bazujących na rozkładzie normalnym (suma, iloczyn, logarytm zmiennych o rozkładzie normalnym) i interesujących własnościach (centralne twierdzenie graniczne). Ponieważ przy metodzie największej wiarygodności różniczkowanie iloczynu może być czasami zbyt uciążliwe, stosuje się tzw. logarytmiczną funkcję wiarygodności, a jak wiadomo z podstaw analizy matematycznej, maksima obu funkcji są identyczne. Również z elementarnej statystyki matematycznej wiadomo, że estymatory uzyskane metodą największej wiarygodności są zgodne oraz asymptotycznie nieobciążone, co dodatkowo sprawia, że są one „dobrymi” estymatorami dla wielu typów modeli rozkładu cechy w populacji.

5. Miscellanea

Możemy pokusić się o obliczenie górnej granicy maksymalnej liczby chromosomów uzyskanych za pomocą operatorów „wzmocnienia” lub „osłabienia”. Zakładając, że mamy t epok (t kolejnych generacji) oraz że $1 \leq m_j \leq L \geq 3$ (przyjmujemy, że chromosom ma co najmniej 3 bity – to założenie zostanie wyjaśnione później) maksymalną liczbę $N_{t_{max}}$ chromosomów w danej epoce t możemy wyznaczyć rekurencyjnie, a mianowicie:

$$1. \text{ W pierwszej epoce mamy } N_{1_{max}} = \binom{n}{2} \cdot \sum_{j=1}^L 2^{L-m_j},$$

gdzie $j = 1, 2, \dots, \binom{n}{2}$, n oznacza liczbę osobników w populacji rodziców.

2. W drugiej epoce mamy $N_{2_{max}} = N_{1_{max}} \cdot \sum_{j=1}^L 2^{L-m_j}$.

3. W trzeciej epoce mamy

$$N_{3_{max}} = N_{2_{max}} \cdot \sum_{j=1}^L 2^{L-m_j} = (N_{1_{max}} \cdot \sum_{j=1}^L 2^{L-m_j}) \cdot \sum_{j=1}^L 2^{L-m_j} = N_{1_{max}} \cdot (\sum_{j=1}^L 2^{L-m_j})^2.$$

.....

t. W t epoce mamy $N_{t_{max}} = N_{t-1_{max}} \cdot \sum_{j=1}^L 2^{L-m_j} = N_{1_{max}} \cdot N_{2_{max}} \cdot \dots \cdot N_{1_{max}} \cdot (\sum_{j=1}^L 2^{L-m_j})^{t-1}$.

Obliczmy ostatnie wyrażenie. Mamy:

$$\begin{aligned} N_{t_{max}} &= N_{1_{max}} \cdot (\sum_{j=1}^L 2^{L-m_j})^{t-1} = \binom{n}{2} \cdot (\sum_{j=1}^L 2^{L-m_j})^t = \frac{(n-1)!}{2} \cdot (\sum_{j=1}^L 2^{L-m_j})^t \\ &= \frac{(n-1)n}{2} \cdot (1 + 2 + 2^2 + 2^3 + \dots + 2^L)^t = \frac{(n-1)n \cdot 2^{(L-1)t}}{2} = \frac{(n-1)n \cdot 2^{tL-t}}{2} = (n-1) \cdot n \cdot 2^{tL-t-1}. \end{aligned}$$

Założenie, że $L \geq 3$ wynika z faktu, iż chcemy, aby na chromosomach były możliwe do wykonania wszystkie zdefiniowane wcześniej operacje genetyczne. Jest oczywistą rzeczą, że użycie wszystkich zdefiniowanych operatorów znacznie poszerza przestrzeń przeszukiwania chromosomów.

Jak widać z otrzymanego wzoru, możemy poprzez dobór parametrów: n (liczność osobników w populacji rodzicielskiej), t (liczba kolejnych generacji pokoleń) oraz dobór operatorów genetycznych definiować żadaną wielkość próbki ankietarskiej.

Modułem, który ma duży wpływ na przebieg pracy algorytmu genetycznego, jest moduł selekcji chromosomów. Jedną z metod wspomnianych w artykule jest metoda selekcji rankingowej. Osobnikom populacji przypisane są pewne rangi w uporządkowanej liście, która określa ich miejsce. Lista ta definiuje stopień przystosowania danego chromosomu od najgorszego do najlepszego lub odwrotnie. Zaletą tej metody jest wprowadzenie do kolejnej populacji pożądaných przez nas osobników, mających dobrą „przystosowalność”. Liczba kopii danego osobnika zależy od funkcji, która z kolei zależy od jego rangi. Mając teraz zbiór Z chromosomów wykorzystujący ten rodzaj selekcji, rozważmy rodzinę wszystkich jego podzbiorów $R \subseteq 2^Z$ spełniających następujące dwa warunki:

w1) $\emptyset \in R$,

w2) jeśli $B \subseteq A \wedge A \in R$, to $B \in R$.

Wtedy parę $P = (Z, R)$ nazywamy systemem niezależnym, elementy rodziny R elementami niezależnymi, a elementy $2^Z \setminus R$ elementami zależnymi systemu P, przy czym symbol „\” oznacza różnicę zbiorów.

Jeżeli dodatkowo jest spełniony warunek:

w3) dla dowolnych A i B $\in R$ takich, że $|B| = |A| + 1$,

to istnieje element $e \in B \setminus A$ taki, że $A \cup \{e\} \in R$.

Wtedy system $P=(Z,R)$ nazywamy matroidem i oznaczamy $M=(Z, R)$. Element $\{e\}$ utożsamiany jest z pojedynczym chromosomem.

Jeżeli teraz postawimy problem P_{\max} maksymalizacji rang chromosomów $r(ch_i)$ na pewnym podzbiory A ($i=1,2, \dots, |A|$), czyli interesuje nas wartość maksymalna przebiegająca zbioru A rodziny \mathcal{R} , to musimy rozwiązać następujący, poniżej zdefiniowany, problem optymalizacyjny – (maximum rang):

$\max_{A \in \mathcal{R}} r(A) = \max_{A \in \mathcal{R}} \sum_{ch \in A} r(ch)$, gdzie $r(A)$ jest sumą rang zbioru chromosomów należących do zbioru A , przy czym zbiór Z stanowi uporządkowaną listę chromosomów LST o mocy $|LST| = n$.

Jeżeli uporządkujemy chromosomy według wzrastającej wartości rang (funkcja $f: Z \rightarrow R^+$, nieujemna funkcja przyporządkowująca każdemu chromosomowi ch_A jego rangę $r(ch_A) \geq 0$), to możemy dla rozwiązania tego problemu zastosować algorytm zachłanny dla matroidu, który posiada własność „optymalnego rozwiązania” – w tym przypadku poszukiwanego maksimum sum rang zbioru chromosomów. Prawdziwe jest następujące twierdzenie, które podajemy tutaj bez dowodu (prosty dowód można znaleźć np. w pracy [4]).

TWIERDZENIE

Jeśli $M=(Z, \mathcal{R})$ jest matroidem, to zbiór A^ znaleziony przez algorytm zachłanny jest optymalnym rozwiązaniem P_{\max} . Jeśli natomiast system $M=(Z, \mathcal{R})$ nie jest matroidem, to zbiór A^* nie jest zbiorem niezależnym o największej sumie rang elementów tego zbioru.*

Algorytm zachłanny dla matroidu jest bardzo prosty. Główna jego idea wygląda następująco:

Krok 1.

$A = \emptyset$ (start algorytmu)

Krok 2. (iteracyjny)

Dla $i = 1, 2, \dots, |Z|=|LST|=n$ wykonaj następującą operację:

jeżeli $A \cup \{e_i\} \in \mathcal{R}$ to $A := A \cup \{e_i\}$.

Krok 3.

Zsumuj rangi zbioru elementów A , tj. oblicz $r(A)$.

Krok 4.

STOP algorytmu.

6. Podsumowanie

Przedstawiona w artykule propozycja rozszerzenia zbioru operatorów genetycznych o dwa dodatkowe, nazwane „wzmocnieniem genowym” i „osłabieniem genowym” ma na celu poprawę wiarygodności otrzymywanych przez ankieterów danych. Z jednej strony zwiększa

to liczbę chromosomów, którym należy się przyjrzeć, gdyż mogą być potencjalnym rozwiązaniem, ale z drugiej strony, poprzez generowanie chromosomów o podobnej strukturze w populacji potomnej, zwiększa się prawdopodobieństwo uzyskania suboptymalnego, a w niektórych przypadkach optymalnego rozwiązania. W pracy przedstawiono również algorytm eliminacji zdublowanych chromosomów, co ma wpływ na złożoność obliczeniową otrzymania wyniku. Następnym krokiem powinno być sprawdzenie i potwierdzenie w praktyce badań ankietowych prezentowanych w artykule idei zastosowania algorytmów genetycznych w badaniach ankietowych z wykorzystaniem dedykowanego oprogramowania.

Bibliografia

1. Mazur, Z., Pec, J. (2016). Diagnoza informatyki statystycznej w aspekcie zarządzania bezpieczeństwem informacji. W B. Królikowska, M. Łatuszyńska (red.), *Metody w badaniach naukowych. Wybrane problemy i zastosowania* (s. 57–77). Warszawa: PTI.
2. Oleński, J. (2018). *Strategie rozwoju e-państwa w warunkach globalizacji – studium Foresight 2030*. Warszawa: Uczelnia Łazarskiego (w przygotowaniu – przytoczono cytaty z pracy za zgodą prof. J. Oleńskiego).
3. Rutkowski, L. (2012). *Metody i techniki sztucznej inteligencji* (s. 241). Warszawa: PWN.
4. Walukiewicz, S. (1986). *Analiza systemowa i jej zastosowania. Programowanie dyskretne* (s. 259–260). Warszawa: PWN.

