

JOANNA HOLDYS
SŁAWOMIR LECIEJEWSKI

Transfer idei z biologii do informatyki na przykładzie algorytmów genetycznych

Wprowadzenie

Niewątpliwie jedną z ważnych i szybko rozwijających się dyscyplin naukowych jest współczesna informatyka. Jej najnowsza historia zaczęła się w pierwszej połowie XX wieku, w latach trzydziestych i czterdziestych podjęto bowiem udane próby skonstruowania urządzeń liczących z zastosowaniem układów elektromechanicznych (była to tak zwana zerowa generacja komputerów). Pierwszą generację komputerów lampowych zapoczątkował uruchomiony w 1945 roku ENIAC. Niestety, ich istotnym mankamentem była wysoka awaryjność lamp zużywających ogromne ilości energii elektrycznej oraz ich bardzo duże gabaryty. Wynalezienie w 1947 roku tranzystora, półprzewodnikowego elementu zastępującego lampy elektronowe, i uruchomienie w 1951 roku pierwszego komputera tranzystorowego wyeliminowało te wady komputerów lampowych. W ten sposób powstała druga generacja komputerów [Gawrysiak, 2008, s. 124–149].

Zmniejszenie odległości między częściami składowymi komputera spowodowało zwiększenie szybkości jego działania. Nadal pozostała jednak istotna niewygodność związana z koniecznością prowadzenia tysięcy połączeń między tranzystorami i innymi elementami komputera. Chęć przewyciężenia tej trudności doprowadziła do powstania w 1958 roku

układu scalonego, a w 1971 roku skonstruowano pierwszy mikroprocesor, co zapoczątkowało pojawienie się trzeciej generacji komputerów konstruowanych na bazie układów scalonych [Gawrysiak, 2008, s. 153–162].

Tak zarysowany rozwój komputerów oraz biegnący równolegle postęp w zakresie oprogramowania i urządzeń peryferyjnych umożliwił wykonywanie w coraz efektywniejszy sposób funkcji obliczeniowych, kontrolnych, doradczych, diagnostycznych, monitorujących, pomiarowych, sterujących i innych. Od początku lat osiemdziesiątych XX wieku komputery zaczęto stosować w niemalże wszystkich obszarach ludzkiej działalności¹. Obecnie jednym z bardzo ważnych ich zastosowań są symulacje komputerowe [Leciejewski, 2013, s. 77–85] odnoszące się do rozwiązywania określonych problemów badawczych, nazywane algorytmami genetycznymi.

Autorzy algorytmów genetycznych utrzymują, że inspiracją do ich powstania były idee zaczerpnięte z biologii ewolucyjnej². Ten dział wiedzy wyłonił się z XIX-wiecznych nauk biologicznych. W ich ramach na podstawie obserwacji przyrody zauważono fakt zmienności biologicznej. Karol Darwin, jako student medycyny, a później podróżnik, obserwował florę, faunę, skamieliny oraz zachowanie zwierząt w walce o przetrwanie. W tej walce, jak zauważył później Darwin, przeżywały organizmy o korzystnych cechach, a eliminacji ulegały jednostki o cechach mniej korzystnych w danym środowisku. Efektem tego była adaptacja do określonych warunków życia lub nawet, w dłuższym okresie, powstanie jednego lub więcej gatun-

¹ Zrodziła się także potrzeba refleksji filozoficznej nad tą dynamicznie rozwijającą się dziedziną wiedzy [Murawski, 2014], przyczynkiem do której jest niniejszy artykuł. Jego cele i strukturę scharakteryzowano w następnym przypisie.

² Zagadnienie to zostanie szczegółowo przeanalizowane w dalszej części artykułu, gdzie prześledzimy, czy ten deklarowany przez programistów zajmujących się algorytmami genetycznymi transfer idei z biologii do informatyki faktycznie ma miejsce. Jednakże przed omówieniem relacji zachodzących pomiędzy biologią ewolucyjną a algorytmami genetycznymi przypomnimy główne idee biologii ewolucyjnej oraz sposób działania klasycznego algorytmu genetycznego. Ułatwi to lepsze zrozumienie relacji zachodzących pomiędzy biologią ewolucyjną oraz algorytmami genetycznymi (w wypadku niniejszego artykułu, dla prostszego ukazania wzmiankowanych relacji, porównanie to zawężone zostanie tylko do klasycznego algorytmu genetycznego). Warto także nadmienić, że niniejszy artykuł stanowi jedynie składową w dyskusji nad transferem wiedzy i metod między dwoma dziedzinami nauki: biologią i stosowanymi naukami komputerowymi. Całościowe przedyskutowanie relacji pomiędzy tymi dziedzinami wykracza poza ramy niniejszego opracowania.

ków potomnych. Darwin zebrał swoje obserwacje i argumentacje na rzecz zmian gatunkowego składu świata żywego w przełomowym dla nauki dziele *O powstawaniu gatunków drogą doboru naturalnego, czyli o utrzymywaniu się doskonalszych ras w walce o byt*, które ukazało się w 1859 roku.

Główne idee biologii ewolucyjnej

Cała Darwinowska teoria ewolucji zasadzała się na zaobserwowanych w przyrodzie czterech składowych mechanizmu doboru naturalnego. Pierwsza z nich to nadprodukcja, czyli wydawanie na świat większej liczby potomstwa niż było w stanie osiągnąć wiek reprodukcyjny, co wydatnie zwiększało szanse przetrwania gatunku. Drugą składową tego mechanizmu stanowiła zmienność międzyosobnicza. Naturalnie występujące różnice między osobnikami w populacji oraz zjawisko podobieństwa dzieci do rodziców Darwin intuicyjnie przypisywał dziedziczności, ponieważ nie sformułowano jeszcze podstaw genetyki. Trzeci mechanizm doboru naturalnego – współzawodnictwo – zakładał, iż osobniki, aby przeżyć, konkurują o pokarm, zasoby środowiska, siedlisko. Celem walki o byt był ostatni z zaobserwowanych mechanizmów, czyli przetrwanie osobników o najkorzystniejszych cechach oraz ich powielenie poprzez prokreację. Zatem darwinowski dobór naturalny prowadził do zwiększenia proporcji udziału korzystnych zmian (cech, genów) w populacji, a zmniejszenie, a w skrajnym wypadku nawet eliminację mniej „pożądanych”, co prowadziło do lepszego przystosowania gatunku do panujących warunków środowiskowych. Kumulacja tychże zmian w czasie i w populacjach odległych geograficznie prowadziła do powstawania nowych gatunków.

Z punktu widzenia współczesnej nauki wyżej omówiona teoria ewolucji Darwina jest niepełna. Darwin nie wiedział bowiem, co jest źródłem różnic międzyosobniczych oraz jak cechy są przekazywane następnym pokoleniom, choć obserwował oba te zjawiska w populacji. Nie znał bowiem podstaw genetyki dziedziczenia, opracowanych przez Grzegorza Mendla w 1865 roku. W latach pięćdziesiątych ubiegłego wieku połączono

teorię ewolucji Darwina z wiedzą, jakiej dostarczył rozwój genetyki Mendla oraz genetyki populacyjnej, tworząc syntetyczną teorię ewolucji. Rozbudowano w niej komponentę zmienności genetycznej w doborze naturalnym o mutacje i rekombinacje, wyjaśniając tym samym mechanizm zmienności międzypersonalnej.

Przymiotnik „genetyczny” jest obecnie bardzo często używany. Genetyka (gr. *genesis* – pochodzenie) dzięki intensywnemu rozwojowi technologicznemu również bardzo dynamicznie się rozwija. Analizy są przeprowadzane na poziomie molekularnym, co z kolei owocuje aplikacją wiedzy teoretycznej w praktyce, między innymi w nanotechnologii.

Jednym z podstawowych pojęć genetyki jest gen, czyli najmniejsza jednostka funkcjonalna, która koduje informację o cechach organizmu, podlegająca dziedziczeniu. Molekularnie gen jest fragmentem (lub zestawem fragmentów) nici DNA (kwasu deoksyrybonukleinowego)³, którego najmniejszą „cegiełkę” stanowi nukleotyd. Cały zapis genetyczny danego organizmu tworzy sekwencja połączonych ze sobą w różnej kolejności czterech rodzajów nukleotydów (A, G, C, T), powtarzających się miliardy razy w DNA. Pomimo że nic DNA składa się zaledwie z czterech rodzajów nukleotydów, to różna kolejność ich ułożenia umożliwia zapisanie dowolnej informacji i stanowi o unikatowości każdego organizmu.

Kilkanaście lub więcej genów leżących na niezwykle pozwijanej nici DNA⁴ tworzy większą jednostkę zwaną chromosomem, natomiast zestaw chromosomów, w charakterystycznej liczbie i kształcie dla danego gatunku, stanowi jego genotyp i zawiera pełny genom organizmu. Na przykład genotyp człowieka zawiera 46 chromosomów (23 pary) w jądrze komórkowym każdej komórki ciała. Unikalny dla danego organizmu zestaw genów to genotyp, a jego widzialnym przejawem jest fenotyp, czyli zespół obserwowanych cech organizmu [Solomon i in., 1996, s. 211].

Wszelkie zmiany w zapisie informacji genetycznej w DNA, czyli mutacje, są kluczowe w procesie ewolucji, ponieważ prowadzą do zwiększenia różnorodności puli genów w populacji. Mogą mieć one charakter

³ Warto zauważyć, że zapis informacji dziedzicznej w przypadku wielu wirusów dokonuje się przy udziale kwasu rybonukleinowego (RNA).

⁴ Od kilkudziesięciu (chromosom X) do ok. 2000 (chromosom 1) u człowieka.

zmian punktowych, czyli wymiany pojedynczego nukleotydu, rearanżacji na poziomie genów chromosomów, na przykład ubytek lub dodatkowa kopia fragmentu chromosomu, lub też mogą dotyczyć całego genomu, czyli zmian ploidalności organizmu (liczby chromosomów) [Voordeckers, Verstrepen, 2015]. Mutacje są zdarzeniami przypadkowymi i spontanicznymi. Nie stanowią wystarczającej przyczyny ewolucji, lecz wraz z mającymi zdecydowanie szerszy zasięg rekombinacjami są podstawowym źródłem zmienności. Występują pod wpływem różnych czynników, na przykład fizycznych, typu promieniowanie UV, chemicznych lub biologicznych, na przykład spadku sprawności enzymów syntetyzujących DNA, na każdym etapie powstawania i rozwoju organizmu. Mutacje występują losowo, to znaczy, że nie zawsze dopasowują się do procesu adaptacji. Prawdopodobieństwo wystąpienia mutacji korzystnej ewolucyjnie jest takie samo jak mutacji obojętnej lub niekorzystnej w procesie ewolucji danej populacji. Jednakże to nie mutacje, ale nowy fenotyp podlega weryfikacji adaptacyjnej. Tempo mutacji jest względnie stałe i nie wszystkie zmiany są dziedziczone. Współczesna biologia molekularna wprowadziła podział zmian materiału dziedzicznego na mutacje, czyli zmiany najczęściej letalne lub wywołujące różne schorzenia, oraz tak zwane polimorfizmy – zmiany zwiększające różnorodność cechy bez znamion choroby, na przykład różne kolory oczu. Przyjmuje się, że mutacje zdarzają się z częstością nie większą niż 1%. Jeśli dana zmiana w materiale genetycznym występuje u więcej niż 1% populacji to mamy do czynienia z polimorfizmem. Z reguły nie wywołują one tak wielkiego efektu fenotypowego jak mutacje i często mogą być dziedziczone bez zauważenia [Słomski, 2011].

Naturalnie występującymi procesami, które mają udział w zmienności genetycznej, są: rekombinacja (*crossing-over*), czyli wymiana fragmentów chromosomów w dzielących się komórkach, oraz transfer genów, polegający na przekazaniu części genów z jednego organizmu do drugiego. Rekombinacja zachodzi zawsze w procesie mejozy, natomiast transfer genów naturalnie występuje u organizmów niższych, na przykład bakterii [Barton, 2010].

Widać zatem, że dziedziczenie odbywa się według pewnych zasad, ale sam proces nie jest bezbłędny, co skutkuje większą różnorodnością i sta-

nowi siłę napędową ewolucji. Ograniczone zasoby środowiska sprawiają, że w walce o byt organizmy podlegają selekcji. Te lepiej przystosowane do warunków środowiska, które zgromadziły więcej korzystnych zmian genetycznych, są faworyzowane przez dobór naturalny, mają większe szanse przeżycia i wydania potomstwa, co prowadzi do utrwalenia korzystnego genotypu w populacji. Natomiast mutanty reprezentujące skrajne cechy są eliminowane z populacji, co zapewnia stabilizację populacji.

Mutacja działa spontanicznie i wielokierunkowo na genotyp, natomiast dobór naturalny działa koordynująco na fenotyp organizmów, porządkuje zmienność genetyczną, eliminując fenotypy, które nie mogą sprostać wyzwaniom środowiska, a wspierając przekazywanie genów przez organizmy najlepiej przystosowane.

We względnie stałym środowisku dobór naturalny będzie eliminował fenotypy skrajne, faworyzując fenotypy przeciętne i zawężając zakres zmienności populacji. Zjawisko to nosi nazwę selekcji stabilizującej. Większość populacji przez dłuższy czas podlega działaniu tej właśnie postaci selekcji. Niektóre gatunki, na przykład miłorząb, posiadają genotyp tak dobrze wyselekcjonowany i dostosowany do środowiska, że ich fenotyp nie podlega zmianom od około 200 milionów lat. W zmieniających się warunkach środowiska zachodzi jednak selekcja kierunkowa, kiedy faworyzowany jeden ze skrajnych fenotypów stopniowo będzie wypierał pozostałe. Szczególnym przypadkiem selekcji kierunkowej jest selekcja rozdzielająca, która zachodzi przy drastycznych zmianach środowiska i polega na działaniu wielokierunkowym. Fenotypy przeciętne są eliminowane na korzyść fenotypów skrajnych, powodując powstawanie różnych grup osobników w danej populacji [Snell-Rood, 2012; Solomon, 1996, s. 419–422].

Na proces ewolucji mają również wpływ inne zjawiska, jak na przykład dryf genetyczny. Ma on miejsce wtedy, kiedy okoliczności tak się układają (katastrofy biologiczne, izolacja lub migracje części populacji na nowy obszar), że krzyżują się ze sobą osobniki w małej liczebnie populacji. Prowadzić to może do eliminacji niektórych wariantów genów z puli genowej całej populacji i w konsekwencji do zmniejszenia zmienności genetycznej i fenotypowej. Ograniczenie puli genów ma miejsce także

wtedy, kiedy populacja przechodzi przez genetyczne wąskie gardło – została zdziesiątkowana i odradza się na nowo z ograniczonej puli genów niewielu organizmów, które przeżyły. Podobna sytuacja ma miejsce, kiedy co najmniej jeden organizm skolonizuje nowe siedlisko, stając się założycielem. Ponieważ przynosi ze sobą niewielką część puli genowej całej populacji, to założona przez niego linia może znacząco różnić się od pozostałych linii tego samego gatunku. Mniejsza zmienność genetyczna, spowodowana izolacją terytorialną, oznacza często mniejsze zdolności adaptacyjne i jest ewolucyjnie niekorzystna, dlatego jest równoważona przez zjawisko migracji, które ułatwia przepływ genów pomiędzy różnymi grupami organizmów tego samego gatunku. Obniżając poziom zmienności pomiędzy populacjami, mechanizm ten przeciwdziała skutkom dryfu genetycznego i doboru naturalnego. Około 300 lat temu zakończył się długi okres względnej izolacji populacji ludzkich zamieszkujących różne kontynenty. Przepływ genów pomiędzy nimi niweluje różnice fenotypowe i powiększa pulę genową każdej z tych populacji, której migracja dotyczy. Warto podkreślić, że im większą pulę genową posiada populacja, tym lepsze wykazuje zdolności adaptacyjne do środowiska [Solomon, 1996, s. 419–422].

Efektom działania procesu ewolucji, a szczególnie doboru naturalnego, nie jest zatem powstanie organizmu „doskonałego”, a jedynie sprzyjanie harmonijnemu współdziałaniu zmienności genetycznej i środowiska, co prowadzi do adaptacji. Wszystkie opisane mechanizmy, choć bywają losowe i nieukierunkowane, a czasem wręcz działające niekorzystnie dla ewolucyjnego rozwoju, są wprzęgnięte w dążenie do osiągnięcia wartości optymalnych w danym miejscu i czasie, zapewniając elastyczność i *continuum* samemu procesowi ewolucji.

Klasyczny algorytm genetyczny

Przypomniane w poprzednim rozdziale podstawowe idee biologii ewolucyjnej stały się inspiracją dla Johna Hollanda, twórcy klasycznego algo-

rytmu genetycznego⁵. Twierdził on bowiem, że przez odpowiednie zaimplementowanie tych idei w formie algorytmu komputerowego otrzymamy nową metodę rozwiązywania złożonych problemów obliczeniowych.

W pierwszym modelu algorytmów genetycznych, który uznawany jest współcześnie za model klasyczny, chromosomy reprezentowane były w formie ciągów zer i jedynek, które kodowały dopuszczalne rozwiązania danego problemu. Algorytmy Hollanda symulowały ewolucję populacji (zbiorów) takich cyfrowych chromosomów, aby znaleźć zadowalające rozwiązanie danego problemu poprzez losowe manipulacje tymi chromosomami. Chromosomy, po każdym cyklu cyfrowej ewolucji, tj. wygenerowaniu zmiany i oceny „przystosowania”, były oceniane pod względem jakości rozwiązania problemu i to decydowało o tym, które z nich będą miały większe, a które mniejsze szanse na dalszą wirtualną reprodukcję.

Pomimo że algorytmy genetyczne Hollanda używały prostej metody kodowania i prostych mechanizmów reprodukcji, to i tak okazały się bardzo skuteczne w rozwiązywaniu wielu złożonych problemów. John Holland w swej monografii *Adaptation in Natural and Artificial Systems* [Holland, 1975] nakreślił podstawowy model, zwany dziś klasycznym algorytmem genetycznym. Zrozumienie ogólnej zasady jego działania wystarczy, aby w kolejnej części artykułu dokonać porównania podstawowych idei biologii ewolucyjnej z ideami ewolucyjnymi zaimplementowanymi w klasycznym algorytmie genetycznym.

Klasyczny algorytm genetyczny operuje na zbiorach łańcuchów złożonych z zer i jedynek. Pojedynczy element takiego zbioru, przyjmujący wartość zero lub jeden, nazywamy genem. Łańcuchy takich zero-jedynkowych genów nazywa się chromosomami⁶, natomiast zbiór cyfrowych

⁵ Algorytmy ewolucyjne to wspólna nazwa kilku podstawowych nurtów rozwojowych w informatyce czerpiących inspiracje z ustaleń biologii ewolucyjnej. Należą do nich: programowanie ewolucyjne, strategie ewolucyjne, algorytmy genetyczne oraz genetyczne programowanie [Arabas, 2004, s. 17–19]. Twórcą jedynie jednego z tych nurtów – algorytmów genetycznych – jest John Holland [Holland, 1975]. Koncepcję tę spopularyzował później David E. Goldberg [Goldberg, 1995].

⁶ W naukach biologicznych występuje dodatkowo pojęcie genotypu, tj. podzbioru chromosomów, które definiują danego osobnika. W klasycznym algorytmie genetycznym przyjęto, że genotyp składa się z pojedynczego chromosomu (jest więc haploidem), zatem

chromosomów o określonej liczebności nazywa się populacją. Holland przyjął, że w ramach określonej symulowanej populacji wszystkie cyfrowe chromosomy są tej samej, ale ściśle określonej długości. Warto dodać, że zarówno długość cyfrowego chromosomu, jak i liczebność symulowanej populacji określane są na etapie projektowania klasycznego algorytmu genetycznego dla rozwiązania konkretnego problemu i mocno zależą od jego charakteru (na przykład inna długość chromosomów i liczebność populacji będzie potrzebna dla rozwiązania problemu komiwojażera⁷, a inna – dla znalezienia wydajnego algorytmu sortującego dane).

Dla tak zdefiniowanej cyfrowej populacji określane są: mechanizm generacji populacji początkowej, mechanizm oceniający jakość konkretnego chromosomu, mechanizm selekcji do dalszego przetwarzania (reprodukcji), mechanizm mutacji oraz mechanizm krzyżowania [Rutkowski, 2011, s. 240–244]. Zostaną one mówione w dalszej części artykułu, gdyż zrozumienie ich działania konieczne jest do uchwycenia specyfiki klasycznego algorytmu genetycznego.

Generacja populacji początkowej polega na losowym utworzeniu żądanej liczby chromosomów (na przykład populacja licząca pięćdziesiąt chromosomów każdy o długości równej dziesięć genów). Najczęściej w tym celu wykorzystuje się tak zwane generatory liczb pseudolosowych, które są częścią każdego systemu komputerowego [Gwiazda, 2007, s. 15–18].

Ocena jakości konkretnego chromosomu polega na obliczeniu wartości tak zwanej funkcji przystosowania. Pokazuje ona, w jakim stopniu chromosom rozwiązuje zadany problem (na przykład jak szybko i dokładnie sortuje zbiór nazwisk). Postać funkcji przystosowania zależy ściśle od charakteru danego problemu i jest jedynym mechanizmem klasycznego algorytmu genetycznego, łącznie z długością chromosomu i liczebnością popu-

w ramach tego modelu każdy osobnik (konkretne rozwiązanie danego problemu) definiowany jest poprzez jeden chromosom [Gwiazda, 1995, s. 10–11].

⁷ Problem komiwojażera polega na znalezieniu najkrótszej drogi pomiędzy n punktami. Problem ten jest rozwiązywalny, ale najprostszy algorytm sprawdza wszystkie n -silnia możliwe drogi i podaje, która z nich jest najkrótsza. Warto jednak podkreślić, że dla n większego od 30 czas rozwiązania problemu komiwojażera jest znacznie dłuższy niż wiek Wszechświata. Algorytmy genetyczne radzą sobie z tym problemem znacznie szybciej [Harel, 2002, s. 92–94].

lacji, każdorazowo zmienianym na etapie projektowania tego algorytmu w celu rozwiązania danego problemu informatycznego. W teorii algorytmów genetycznych zakłada się, że funkcja przystosowania przyjmuje zawsze wartości nieujemne; co więcej, zakłada się także, że rozwiązywany problem jest problemem znalezienia maksimum tej funkcji [Gwiazda, 2007, s. 18–20].

Selekcja chromosomów do dalszego przetwarzania polega na wybraniu zbioru chromosomów o tej samej liczebności co populacja początkowa, które staną się cyfrowymi „rodzicami” nowo tworzonej populacji potomków. Innymi słowy, chodzi o wybranie zbioru przeznaczanego do wirtualnej reprodukcji. Warto podkreślić, że selekcja ta ma przebieg losowy, jednakże przeprowadzana jest w taki sposób, aby chromosomy o największej wartości funkcji przystosowania miały największe szanse być wylosowanymi do reprodukcji (może się tak zdarzyć, że chromosomy o największej wartości funkcji przystosowania wylosowane zostaną wielokrotnie). Następnie wylosowane do reprodukcji chromosomy łączone są losowo w pary cyfrowych rodziców, które na drodze wirtualnego krzyżowania i wirtualnej mutacji przekształcone zostaną w pary potomków, tworząc tym samym nową cyfrową populację [Gwiazda, 2007, s. 21–23, 27–31].

W klasycznym algorytmie genetycznym mutacja chromosomu jest mechanizmem oddziaływującym na populację rodziców przed procesem krzyżowania lub na populację potomków utworzoną w wyniku krzyżowania. Cyfrowa mutacja zaczyna się od ustalenia, na etapie projektowania klasycznego algorytmu genetycznego, prawdopodobieństwa zaistnienia mutacji (zakłada się, że jest ono bardzo małe). Następnie dla każdego genu w każdym chromosomie populacji losowana jest liczba, która jeśli okaże się mniejsza lub równa wcześniej ustalonemu prawdopodobieństwu zaistnienia mutacji, doprowadza do mutacji danego genu. Mutacja taka prowadzi do zmiany wartości genu, na przykład z zera na jeden lub odwrotnie. W klasycznym algorytmie genetycznym rolą mutacji jest nie tyle zmiana jakościowa chromosomów, co wprowadzenie nowej informacji (nowego cyfrowego materiału genetycznego) do danej populacji [Gwiazda, 2007, s. 23–27].

Krzyżowanie par cyfrowych chromosomów ma na celu wymianę części cyfrowego materiału genetycznego pomiędzy symulowanymi chromosomami, co może zaowocować wirtualnym potomkiem dziedziczącym na przykład tylko lepsze cechy swoich wirtualnych rodziców. Przebieg krzyżowania projektowany jest na etapie powstawania klasycznego algorytmu genetycznego, kiedy to ustala się prawdopodobieństwo krzyżowania. Zakłada się, że jest ono stosunkowo duże. Następnie dla każdego chromosomu z populacji rodziców losowana jest liczba, która jeśli jest mniejsza lub równa prawdopodobieństwu krzyżowania, to dany chromosom podlega krzyżowaniu. Z grupy chromosomów przeznaczonych do krzyżowania tworzone są losowo pary rodziców (w przypadku, gdy liczebność tej grupy jest nieparzysta, losuje się jeden dodatkowy chromosom do zbioru rodziców). Dla każdej pary rodziców losuje się także punkt krzyżowania, czyli numer genu n , w miejscu którego nastąpi krzyżowanie. Potomek pierwszy otrzymuje n genów od rodzica pierwszego, a pozostałe geny, począwszy od genu $n+1$, od rodzica drugiego. Drugi potomek otrzymuje pierwsze n genów od rodzica drugiego a pozostałe geny od rodzica pierwszego [Gwiazda, 2007, s. 23–27].

Mając zdefiniowane wszystkie mechanizmy oddziałujące na cyfrową populację chromosomów w toku symulowanej ewolucji, warto przedstawić schemat działania klasycznego algorytmu genetycznego [Rutkowski, 2011, s. 240–252]:

- etap 0: generacja populacji początkowej,
- etap 1: obliczanie wartości funkcji przystosowania chromosomów należących do populacji początkowej,
- etap 2: selekcja, za pomocą metody losowej, cyfrowych chromosomów przeznaczonych do reprodukcji,
- etap 3: tworzenie, za pomocą operatorów mutacji i krzyżowania, nowej populacji potomków,
- etap 4: obliczanie wartości funkcji przystosowania chromosomów należących do populacji potomków,
- etap 5: populacja rodziców jest kasowana; populacja potomków staje się nową populacją rodziców,

- etap 6: warunek końca (warunek stopu) działania algorytmu (na przykład poprawne alfabetyczne posortowanie miliona nazwisk w czasie jednej minuty):
- jeśli jest spełniony, wtedy kończymy wykonywanie algorytmu genetycznego (rozwiązaniem danego problemu numerycznego staje się chromosom kodujący rozwiązanie spełniające warunek końca działania algorytmu),
- jeśli nie jest spełniony, wtedy powracamy do etapu 2.

Warto dodać, że warunek końca (warunek stopu) działania algorytmu ustalany jest na etapie projektowania klasycznego algorytmu genetycznego i w najprostszym przypadku jest konsekwencją ograniczeń czasowych (na przykład zakłada się z góry określoną liczbę iteracji – skoków z etapu 6 do etapu 2). Najlepszy chromosom z ostatniej cyfrowej populacji potomków traktowany jest jako rozwiązanie danego problemu⁸. Nie ma jednak gwarancji, że jest to rozwiązanie najlepsze z możliwych (rozwiązanie optymalne) [Gwiazda, 1995, s. 18–21].

Biologia ewolucyjna a algorytmy genetyczne

„W ciągu ostatnich 30 lat narastało zainteresowanie systemami, w których do rozwiązywania zadań stosuje się zasady ewolucji i dziedziczności. W systemach tych występuje populacja potencjalnych rozwiązań, zawierają one pewien proces selekcji, oparty na dopasowaniu osobników, i pewne operatory «genetyczne». Jednym z typów takich systemów jest klasa strategii ewolucyjnych, to znaczy algorytmów, które naśladują zasady ewolucji w naturze przy rozwiązywaniu zadań optymalizacji parametrycznej [...]. Innym typem systemów używających zasad ewolucji są algorytmy genetyczne Hollanda” [Michalewicz, 1999, s. 25]. Warto prześledzić, czy ten deklarowany przez zajmujących się algorytmami genetycznymi infor-

⁸ „Cykl ewolucji może kończyć się wówczas, gdy przystosowanie osobników jest odpowiednio duże, lub gdy stwierdzi się, że stan populacji bazowej świadczy o stagnacji algorytmu” [Arabas, 2004, s. 17].

matyków transfer idei z biologii faktycznie ma miejsce. Jeśli tak jest, to należałoby także odpowiedzieć na pytanie: w jakim zakresie następuje przeniesienie ustaleń z biologii ewolucyjnej do działu informatyki zajmującego się algorytmami genetycznymi?

Już sam przymiotnik „genetyczny”, który współtworzy nazwę omawianego algorytmu, a także składowe procedury i schemat działania zawierają słownictwo zaczerpnięte z języka biologii. Pojęcia takie jak „gen”, „chromosom”, „populacja”, w bardzo dużym uproszczeniu pokrywają się z ich biologicznymi odpowiednikami, choć trudno nie zauważyć różnic pomiędzy biologicznym i informatycznym kontekstem ich użycia.

Alfabet języka cyfrowego, którego używa się do kodowania rozwiązań w ramach klasycznego algorytmu genetycznego, składa się tylko z dwóch liter – jest zero-jedynkowy, a język genetyczny wykorzystuje cztery różne nukleotydy do kodowania informacji. Uproszczeniem opisanych symulacji komputerowych w stosunku do ewolucji biologicznej jest także pojęcie cyfrowego genu, czyli tak naprawdę zero lub jeden w całym ciągu zer i jedynek, który tworzy chromosom. W rzeczywistości biologicznej bowiem gen jest dużo dłuższy niż jedna „cegiełka”.

Schemat działania algorytmu genetycznego opiera się na procesach zachodzących w trakcie ewolucji: mutacji, krzyżowaniu, które w przyrodzie generują zjawisko różnorodności. Inna jest jednak skala i czas, w jakich zachodzą naturalne procesy i te symulowane komputerowo. Jednakże, jak się wydaje, w algorytmach genetycznych nie tyle chodzi o wierne odwzorowanie procesów ewolucyjnych, co jedynie wykorzystanie w konkretnym celu pewnych mechanizmów, które działają w przyrodzie. Stanowią one zatem zespół technik informatycznych, jakie można uznać za zainspirowane przez osiągnięcia biologii, w tym wypadku odnoszące się do mechanizmów rządzących zmiennością form świata istot żywych, dla których utrzymanie zdolności do przetrwania oraz do zwiększenia przewagi konkurencyjnej nad innymi ma znaczenie krytyczne. W tym miejscu dotykamy najistotniejszej różnicy pomiędzy procesami ewolucji biologicznej a jej symulacjami zaproponowanymi w ramach algorytmów genetycznych. Chodzi mianowicie o kwestię celu cyfrowej ewolucji.

Algorytm genetyczny został stworzony z myślą o wypełnieniu konkretnego zadania, rozwiązania danego problemu, przy konkretnych założeniach początkowych i dodatkowych, które dotyczą poszczególnych kroków schematu jego działania [Koza, Keane, Streeter, 2003, s. 40–47]. Celem algorytmu genetycznego jest jak najlepsze rozwiązanie danego problemu, czyli osiągnięcie maksimum pewnej funkcji⁹. Tworząc taki algorytm, z reguły wiemy, jaki wynik chcemy uzyskać, na przykład szybko i dokładnie posortować określony zbiór danych. Możemy nawet założyć, w jakim czasie chcemy mieć owe dane posortowane, i odrzucamy automatycznie wszystkie rozwiązania, które wykonują to zadanie w czasie dłuższym.

W przeciwieństwie do rzeczywistości cyfrowej, proces ewolucji nie jest ukierunkowany. Zmienne warunki środowiska, zdarzenia losowe oraz mechanizmy doboru naturalnego działające na populację prowadzą do zmian w jej puli genowej, a powstające nowe fenotypy są konfrontowane z bieżącymi wymaganiami środowiska. Jeśli przejdą test, mają szansę na włączenie w proces adaptacji. Może się wydawać, że dążenie populacji do jak najlepszego dostosowania się do warunków siedliska (adaptacja) stanowi pewien trend w procesie ewolucji, kształtuje się on jednak spontanicznie i w długim okresie. Sama ewolucja jest procesem ciągłym, wielokierunkowym, losowym, niepowtarzalnym i nieodwracalnym. Poszukiwane jest rozwiązanie optymalne dla danego, choć ciągle zmieniającego się, środowiska. Niekoniecznie musi być to rozwiązanie idealne, najlepsze czy realizujące jakieś maksymalne parametry. Niemożliwe jest zatem określenie celu ewolucji biologicznej dla poszczególnych populacji¹⁰.

⁹ Warto dodać, że „znajdowanie (optymalnego) rozwiązania problemu polega na heurystycznym przeszukiwaniu przestrzeni stanów (w której między innymi znajdują się potencjalne rozwiązania problemu). Główną ideą modeli inspirowanych biologią jest naśladowanie natury w takim procesie przeszukiwania” [Flasiński, 2011, s. 30].

¹⁰ W ściśle określonych warunkach środowiskowych, choć ciągle zmiennych i dlatego trudnych do przewidzenia w przyszłości, można by za quasi-cel ewolucji uznać przystosowanie i przetrwanie oraz przewagę selekcyjną. Warto jednak podkreślić, że środowisko życia jest permanentnie zmienne, zatem i cel takiej ewolucji byłaby ciągle zmienny. W przypadku klasycznego algorytmu genetycznego symulowane środowisko jest niezmienne, zatem cel cyfrowej ewolucji może być precyzyjnie i jednoznacznie zadany przez programistę.

Przykładem próby wyznaczenia końcowego efektu procesu ewolucji z zastosowaniem tylko części kryteriów wchodzących w skład ewolucji naturalnej jest sztuczna selekcja w eugenicie, której celem było przyspieszenie ewolucji człowieka [Klichowski, 2014, s. 21–95]. Zupełnie natomiast niekontrowersyjnym zastosowaniem metod selekcji jest hodowla zwierząt prowadzona w celu uzyskania nowych, lepszych ras, charakteryzujących się konkretnymi, pożądanymi cechami użytkowymi. W kontekście cyfrowej funkcji przystosowania (określającej, jak blisko jesteśmy założonego podczas cyfrowej ewolucji celu), wymienione wyżej eugenika i hodowla są bliskie sposobowi działania algorytmów genetycznych¹¹, jednakże – co warto podkreślić – znacznie odbiegają od podstawowych idei naturalnej ewolucji. W tej ostatniej bowiem nie istnieje coś takiego jak założony z góry cel ewolucji, którego miarą w ramach cyfrowej ewolucji jest funkcja przystosowania.

Tak więc ważną różnicą pomiędzy mechanizmami opisującymi ewolucję w biologii a cyfrową ewolucją w ramach klasycznego algorytmu genetycznego jest ta, która uwidacznia się w nowocześniejszym podejściu do algorytmów genetycznych, uwzględniająca specyfikę rozwiązywanych problemów już na etapie kodowania i operatorów genetycznych [Michalewicz, 1999]. Początkowo tego typu algorytmy

były dość chętnie traktowane jako metody ogólnego przeznaczenia, służące do rozwiązywania dowolnych problemów w identyczny sposób. Wraz z pierwszymi zastosowaniami, na początku lat 90. zaczęto dostrzegać słabość takiego postępowania. Za dużą uniwersalność trzeba mianowicie zapłacić niewielką efektywnością rozwiązywania konkretnego zadania. Pojawiły się propozycje uwzględniania specyfiki zadania poprzez stosowanie specjalizowanego kodowania i operatorów genetycznych [...]. [Arabas, 2004, s. 18–19]

Wynika z tego, że informatycy niezadowoleni z efektywności implementacji ogólnych metod ewolucji w ramach algorytmów genetycznych zaczęli od nich odchodzić w stronę „ręcznego” poprawiania tych mechani-

¹¹ Stąd wydaje się, że lepszą ich nazwą, gdyż wierniej oddającą sposób ich działania (zamiast „algorytmy genetyczne”), byłyby nazwy: „algorytmy eugeniczne” lub „algorytmy hodowlane”.

zmów w zależności od rozwiązywanego problemu [Arabas, 2004, s. 243–266]. Warto jeszcze raz podkreślić, że procesu ewolucji biologicznej nie da się zmieniać w zależności od tego, co i w jakim kierunku miałyby ewoluować. „Ręczne” poprawki w mechanizmach ewolucyjnych dokonuje się tylko w ramach wspomnianych wcześniej hodowli i eugeniki, do symulacji czego – jak się wydaje – zmierzają sposoby realizacji i działania współczesnych algorytmów genetycznych uwzględniające specyfikę rozwiązywanego problemu [Arabas, 2004, s. 243–266].

Kolejną istotną różnicą zachodzącą pomiędzy ewolucją biologiczną a jej symulacją w ramach algorytmów genetycznych jest ta, która odnosi się do warunku końca działania omawianego algorytmu. Algorytm genetyczny kończy działanie, gdy uzyskuje się zadowalające rozwiązanie spełniające założony uprzednio cel cyfrowej ewolucji, którego miarą jest funkcja przystosowania lub po określonej z góry liczbie iteracji (liczbie pokoleń). W naturalnych procesach ewolucji biologicznej nie zachodzi zaden z tego typu czynników kończących ewolucję, która odbywa się w ramach algorytmów genetycznych. Ewolucja nie jest bowiem teleologiczna (o czym była mowa wyżej) i nie zakłada skończonej, z góry ustalonej liczby pokoleń, po zaistnieniu których raptownie się kończy.

Na zakończenie niniejszego rozdziału, w którym dokonano porównania mechanizmów biologii ewolucyjnej z mechanizmami umożliwiającymi działanie algorytmów genetycznych, należy jeszcze raz podkreślić, że takie pojęcia jak „gen”, „populacja”, „selekcja” itd. mają, pomimo tych samych nazw, nieco odmienne znaczenie w ramach symulacji komputerowych i biologii ewolucyjnej. Pojęcia te, występujące w dziale informatyki opracowującym algorytmy genetyczne, zawsze występują w kontekście systemu, który za pomocą funkcji przystosowania zakłada jakiś cel. Teleologiczny kontekst użycia tych pojęć niewątpliwie zmienia ich treść, która w przypadku zastosowania algorytmów genetycznych jest w znacznym stopniu inna niż ta, jaką postulują nauki biologiczne [Mutschler, 2005, s. 143–148].

Za stwierdzeniem, że komputeryzacja przyrody doprowadziła do odnowionej teleologii, opowiada się całkiem bez ogródek John Holland, wynalazca „genetycznych

algorytmów¹². Jego zdaniem, ewolucja jest „pozornie bezsensowna”, wychodzi zatem niewątpliwie od realno-teleologicznie działającej ewolucji, podczas gdy praktycznie wszyscy współcześni biolodzy mówią jedynie o `teleonomicznej`, tzn. o ewolucji *pozornie skierowanej na cel*. [Mutschler, 2005, s. 145–146]

Innymi słowy, informatycy zajmujący się symulowaniem ewolucji w ramach algorytmów genetycznych muszą traktować ją nie teleonomicznie, ale jako realnie skierowaną na cel. W innym wypadku bowiem tego typu wirtualna ewolucja nie będzie mogła się zakończyć. Założony podczas cyfrowej ewolucji cel określa bowiem warunek końca (warunku stopu) programu, na przykład to, że chcę uzyskać poprawne sortowanie określonej liczby nazwisk w wyznaczonym czasie. Bez jasno sprecyzowanego celu (na przykład maksymalnej liczby błędów sortowania i maksymalnego czasu tej operacji) algorytm genetyczny nigdy nie mógłby zakończyć swojej pracy, a tym samym byłby nieużyteczny jak każdy inny program, który zawiera w sobie błędy pętli powtarzających się nieskończoną ilość razy.

Podsumowanie

Algorytmów genetycznych, tj. omówionego wyżej sposobu rozwiązywania problemów optymalizacyjnych i obliczeniowych, można używać w wielu dziedzinach: naukach inżynierskich, edukacji, ekonomii, mediach, naukach empirycznych i wszędzie tam, gdzie nie jest znany efektywny sposób rozwiązania problemu, ale znany jest sposób oceny jakości tego rozwiązania¹². Algorytmy genetyczne mają wiele zalet [Koza, Keane, Streeter, 2003, s. 40–47]. Metoda jest szybka (widać to na przykładzie rozwiązywania problemu komiwojażera), uniwersalna (można użyć tego samego programu do rozwiązywania innego problemu, zmieniając tylko funkcję przystosowania), powtarzalna w celu uzyskania lepszych wyników

¹² Obszerną listę opatentowanych wynalazków ponownie „wymyślonych” przez komputer oraz inne osiągnięcia programowania genetycznego znaleźć można w artykule *O dokonywaniu wynalazków drogą ewolucji* [Koza, Keane, Streeter, 2003, s. 45–47].

(algorytm genetyczny jest algorytmem randomizowanym, zatem możemy wielokrotnie powtarzać obliczenia w nadziei otrzymania lepszych wyników) [Leciejewski, 2014].

Algorytmy genetyczne, oprócz wielu zalet, mają także swoje wady i ograniczenia epistemologiczne. Wydaje się, że najpoważniejszą jest fakt, że aby uzyskać wynik, należy precyzyjnie określić funkcję przystosowania, a nie ma żadnej jednoznacznie określonej procedury, która informowałaby, jak to zrobić. Po drugie, nigdy nie mamy pewności, że rozwiązanie problemu jest optymalne, zwykle bowiem rezultatem działania algorytmu genetycznego jest uzyskanie rozwiązania przybliżonego, gdyż jest to algorytm randomizowany.

Warto na zakończenie jeszcze raz mocno podkreślić, że algorytmy genetyczne w zamierzeniach ich twórców korzystają z ustaleń teorii ewolucji, ale jest między nimi jedna podstawowa różnica: ewolucja biologiczna jest ateleologiczna, a algorytmy genetyczne są teleologiczne. Bez jednoznacznie określonego celu procedury optymalizacyjne, jakimi są algorytmy genetyczne, nie są w stanie konkluzywnie zakończyć swojego działania. W przypadku natomiast ewolucji biologicznej możemy co najwyżej mówić o teleonomii, tzn. ewolucji pozornie skierowanej na cel.

Niewątpliwie zatem ustalenia biologii ewolucyjnej były inspiracją do powstania algorytmów genetycznych, a to oznacza, że miał miejsce transfer idei z biologii do informatyki. Jednakże był to tylko transfer idei wybranych. Cała bowiem „maszyneria `genetycznych algorytmów` pozwala się [...] tylko wtedy użyć, gdy `funkcję *fitness*` potraktujemy jako coś zastępującego jakiś cel. W przeciwnym razie nie można zrozumieć wywołanego przez nią przebiegu optymalizacji” [Mutschler, 2005, s. 146]. Innymi słowy między procesem ewolucji biologicznej a procesami cyfrowej ewolucji w ramach algorytmów genetycznych możemy co najwyżej mówić o pewnych analogiach. Procesy cyfrowej ewolucji zachodzą dzięki „operacjom analogicznym do tych, jakie mają miejsce w przypadku organizmów żywych, na przykład: mutacjom oraz/lub krzyżowaniu rozwiązań-osobników” [Flasiński, 2011, s. 30].

Bibliografia

- Arabas J. (2004), *Wykłady z algorytmów ewolucyjnych*, Warszawa, Wydawnictwa Naukowo-Techniczne.
- Barton N.H. (2010), "Genetic linkage and natural selection", *Philos Trans R Soc Lond B Biol Sci.* 365 (1552), s. 2559–2569.
- Flasiński M., (2011), *Wstęp do sztucznej inteligencji*, Warszawa, Wydawnictwo Naukowe PWN.
- Gawrysiak P., (2008), *Cyfrowa rewolucja. Rozwój cywilizacji informacyjnej*, Warszawa, Wydawnictwo Naukowe PWN.
- Goldberg D.E., (1995), *Algorytmy genetyczne i ich zastosowania*, Warszawa, Wydawnictwa Naukowo-Techniczne.
- Gwiazda T.M., (1995), *Algorytmy genetyczne. Wstęp do teorii*, Warszawa, T.D.G.
- Gwiazda T.M., (2007), *Algorytmy genetyczne. Kompendium. TOM I. Operator krzyżowania dla problemów numerycznych*, Warszawa, Wydawnictwo Naukowe PWN.
- Harel D., (2002), *Komputery – spółka z o.o. Czego komputery naprawdę nie umieją robić*, Warszawa, Wydawnictwa Naukowo-Techniczne.
- Holland J., (1975), *Adaptation in Natural and Artificial Systems*, Ann Arbor, University of Michigan Press.
- Klichowski M., (2014), *Narodziny cyborgizacji. Nowa eugenika, transhumanizm i zmierzch edukacji*, Poznań, Wydawnictwo Naukowe UAM.
- Koza J.R., Keane M.A., Streeter M.J., (2003), „O dokonywaniu wynalazków drogą ewolucyjną”, *Świat Nauki* 4(140), s. 40–47.
- Leciejewski S., (2013), *Cyfrowa rewolucja w badaniach eksperymentalnych. Studium metodologiczno-filozoficzne*, Poznań, Wydawnictwo Naukowe UAM.
- Leciejewski S., (2014), „Ewolucyjna teoria epistemologiczna metodą algorytmów genetycznych”, *Przegląd Filozoficzny – Nowa Seria*, 4(92), s. 263–273.
- Michalewicz Z., (1999), *Algorytmy genetyczne + struktury danych = programy ewolucyjne*, Warszawa, Wydawnictwa Naukowo-Techniczne.
- Murawski R., (2014), *Filozofia informatyki. Antologia*, Poznań, Wydawnictwo Naukowe UAM.
- Mutschler H.D., (2005), *Wprowadzenie do filozofii przyrody*, Kraków, WAM.
- Rutkowski L. (2011), *Metody i techniki sztucznej inteligencji*, Warszawa, Wydawnictwo Naukowe PWN.
- Snell-Rood E.C. (2012), "Selective processes in development: implications for the costs and benefits of phenotypic plasticity", *Integrative and Comparative Biology*; 52(1), s. 31–42.
- Słomski R. (red.), (2011), *Analizy DNA. Teoria i praktyka*, Poznań, Wydawnictwo Uniwersytetu Przyrodniczego w Poznaniu.
- Solomon E.P., Berg L.R., Martin D.W., Villee C.A., (1996), *Biologia*, Warszawa, MULTICO.

Voordeckers K., Verstrepen K.J. (2015), “Experimental evolution of the model eukaryote *Saccharomyces cerevisiae* yields insight into the molecular mechanisms underlying adaptation”, *Current Opinion in Microbiology*; 28, s. 1–9.

Transfer of ideas from biology to computer science based on the example of genetic algorithms

ABSTRACT. Those who have developed genetic algorithms assert that they were inspired by ideas in evolutionary biology. The author of this article investigates whether programmers who deal with genetic algorithms really did borrow ideas from biology and bring them to computer science. In order to facilitate the understanding of these issues, the author also outlines the major ideas in evolutionary biology and the way in which classical genetic algorithms work

KEY WORDS: philosophy of computer science, evolutionary science, algorithms, artificial intelligence

Joanna Holdys, Akademia Wychowania Fizycznego im. Eugeniusza Piaseckiego w Poznaniu, ul. Królowej Jadwigi 27/39, 61-871 Poznań

Sławomir Leciejewski, Instytut Filozofii, Uniwersytet im. Adama Mickiewicza, ul. Szamarzewskiego 89C, 60-568 Poznań, slaaw@amu.edu.pl