

ANDRZEJ GECOW

*Instytut Paleobiologii PAN
Twarda 51/55, 00-818 Warszawa,
E-mail: gecow@twarda.pan.pl*

ALGORYTMY EWOLUCYJNE I GENETYCZNE, EWOLUCJA SIECI ZŁOŻONYCH I MODELE REGULACJI GENOWEJ A MECHANIZM DARWINOWSKI

WSTĘP

Niniejszy zbiór artykułów powstał z okazji 200-nej rocznicy urodzin Karola Darwina i 150-tej rocznicy skutecznego opublikowania przez niego mechanizmu ewolucji biologicznej. Pomińmy tu, że autorstwo tego mechanizmu dzieli on bez oporu z Alfredem Wallacem i potraktujmy Darwina jako przedstawiciela obu odkrywców. Mechanizm darwinowski tłumaczy, skąd bierze się w organizmach żywych widoczna na pierwszy rzut oka wspaniała celowość budowy i funkcjonowania. Zdumienie i zachwyt tą obserwowaną celowością są tak wielkie, że trudno bez głębszego zastanowienia i zrozumienia wskazanego przez Darwina mechanizmu pogodzić się z myślą, że jest ona wynikiem ślepego przypadku. A rzeczywistość jest wynikiem całkiem losowych zmian, tylko że to stwierdzenie to tylko pół prawdy. Drugie pół, to eliminacja – brak przeżywalności (statystyczny) tych organizmów, które mają zbyt małe prawdopodobieństwo przeżycia i pozostawienia potomstwa (zwykle nazywane dostosowaniem, czyli „fitness”¹). W zasadzie jest to tautologia, oczywiste stwierdzenie, ale właśnie ono okazuje się najbardziej podważane. Jak widać, nie ma w tym mechanizmie żadnego z góry założonego „Inteligentnego Planu”, po prostu pozostaje tylko to, co w da-

nych warunkach pozostać może i to właśnie obserwujemy. O tej pozornej, ale jakże ważnej tautologii pisze także ARGASIŃSKI w tym zeszycie KOSMOSU, gdzie pokazuje pewne metody przewidywania „co w danych warunkach pozostać może i dlaczego”.

Próby podważania mechanizmu darwinowskiego trudno dziś zaliczyć do naukowych, zwykle mają podłoże emocjonalne, związane ze skrajnie prostym rozumieniem religii lub godności człowieka. Na XIII Krakowskiej Konferencji Metodologicznej w 2009 r. poświęconej ewolucji, zorganizowanej głównie przez ks. prof. Hellera z Papieskiej Akademii Teologicznej, w podsumowaniu zgodzono się, że nie widać obecnie alternatywy naukowej dla ewolucji darwinowskiej, która dawno przestała być jedynie hipotezą (patrz artykuł BAŃBURY w tym zeszycie KOSMOSU).

Losowa zmienność, nie ograniczona oceną swojego skutku, a więc w kierunkach zwykle gorszych i sporadycznie lepszych (w sensie wzrostu fitness), oraz statystyczna eliminacja tych zmian, które statystycznie pogarszają zdolność „przeżywania” – fitness wystarczy do zbudowania tych wspaniałych, celowych² konstrukcji służących „przeżywa-

¹Fitness jest różnie definiowane. W tym artykule używane jest w ogólnym sensie miernika jakości tak, jak się go używa w optymalizacji a w szczególności w algorytmach ewolucyjnych, ewolucji sieci czy modelach regulacji genowych. Jego interpretacyjny związek z fitness używanym w biologii jako dostosowanie lub przystosowanie nie jest dokładny (patrz też artykuł ŁOMNICKIEGO „Dobór naturalny” i KOZŁOWSKIEGO „Ewolucja strategii życiowych” w tym zeszycie KOSMOSU), dlatego używana będzie na potrzeby tego artykułu nie przetłumaczona nazwa „fitness”.

²Termin „celowych” użyto tu w ogólnym, potocznym znaczeniu. Jest on bardzo kontrowersyjnym terminem, w biologii należałoby stosować bardziej określony termin „teleonomia”. Ma on pokazną literaturę, patrz także GECOW 2008b.

niu". To „przeżywanie” dotyczy pełnego cyklu życiowego, zawiera więc w sobie nie tylko przeżycie osobnika do chwili rozrodu ale i pozostawienie przynajmniej jednego potomka i tak powinno być dalej rozumiane. Fitness jest tu globalną oceną szansy takiego „przeżycia”. Aby to zrozumieć i wewnętrznie pogodzić się z potęgą tego prostego mechanizmu przyjrzymy się mu w jego czystej postaci – Algorytmów Ewolucyjnych. Poznamy najpierw skrajnie prosty przykład by przygotować intuicję. Uświadomimy sobie rolę ograniczonej pojemności środowiska i wynikającej z tego ograniczenia konkurencji.

Konieczność eliminacji i jednoczesnego utrzymania (przez długi czas) przynajmniej jednej nitki ewolucyjnej wymaga rozmnażania. Wystarczy rozmnażanie wegetatywne (z jednego rodzica). Zmienność może zachodzić jednocześnie podczas rozmnażania, ale są to w omawianym mechanizmie osobne wymagania. Oczywiście zmiana ma być trwała poprzez pokolenia, aż nastąpi nowa zmiana tej treści. Tylko taka trwała zmiana nas interesuje – tę cechę nazywamy dziedzicznością. Dla wygody zakłada się zwykle jedną zmianę na pokolenie i definitywną odpowiedź testu w tym czasie – wyznaczenie fitness i ewentualną eliminację. Interpretacja fitness np. jako dostosowanie zawierające pozostawienie przynajmniej jednego potomka, nie jest elementem Algorytmów Ewolucyjnych, jest względem nich zewnętrzna.

Darwin obserwował głównie zwierzęta posiadające płeć i zdolne do rozmnażania generatywnego (dwoje rodziców). Tu też rozdzielmy funkcję rozmnażania od wymiany informacji o budowie i zachowaniu, zwanej krzyżowaniem. Są zresztą organizmy, takie jak orzęski, gdzie procesy te są rzeczywiście oddzielone – namnażanie odbywa się poprzez podział komórek, a wymiana informacji genetycznej poprzez koniugację, gdzie dwie komórki wymieniają się połową materiału genetycznego i dalej pozostają dwoma komórkami, nie ma więc rozmnażania. Podobnie jest nawet u bakterii (WŁODARCZYK 1998). Ponieważ u organizmów wielokomórkowych, po takiej wymianie informacji na podstawie informacji z dwóch różnych źródeł, powstaje zwykle nieco inny od każdego rodzica, nowy konkretny organizm, więc zazwyczaj przy okazji dokonuje się rozmnażanie. Krzyżowanie tworzy realne populacje (zwane mendlowskimi), w ramach których zachodzi owa wymiana. Daje to znacznie silniejszy i szybszy mechanizm ewolucji. Pozwala on na

zbieranie rozmaitych „dobrych pomysłów”, (tj. zmian podwyższających prawdopodobieństwo „przeżycia”), często alternatywnych, które powstały w różnych współistniejących osobnikach. Populacja jest jakby ich magazynem, ale mechanizm ten pozwala na rozmaite ich zestawienia w poszczególnych osobnikach, w tym na utworzenie osobników złożonych ze szczególnie dobrych wariantów. Praktycznie testowane są tu nie całe osobniki, które nadal ulegają eliminacji, a owe „pomysły”, których udział w populacji rośnie lub maleje. W ewolucji biologicznej odbywa się to nawet kosztem gubienia szczególnie dobrych zestawów alternatywnych cech, które wskutek rekombinacji nie przechodzą do następnego pokolenia jako zestaw.

W czystej postaci powyższy mechanizm nosi nazwę Algorytmu Genetycznego (GOLDBERG 1995) i jest chyba najsilniejszym znanym ogólnym sposobem optymalizacji. Kto nie spróbował użyć tego narzędzia do optymalizacji bardzo złożonej funkcji o dziesiątkach parametrów bez potrzeby jej rozwikływania, ten na pewno nie docenił potęgi tego algorytmu. W praktycznych zastosowaniach numerycznych zazwyczaj nie gubi się już znalezionej najlepszego zestawu, tylko czeka na jeszcze lepszy.

Nie jest ważne, czy Darwin wyraźnie rozróżniał powyższe dwa algorytmy (z krzyżowaniem i bez, czyli zazwyczaj z rozmnażaniem płciowym i bez niego) i że nie używał pojęcia „algorytm”. Główna jego idea, to modyfikacja frekwencji cech w populacji po losowej zmianie poprzez eliminację wynikającą z konkurencji. Konkurencja ta wynika z ograniczonych zasobów, które wyznaczają ograniczenie liczebności, a ponieważ rozmnażanie prowadzi do ciągłego wzrostu liczebności, więc pojawia się konieczność eliminacji, oczywiście z większym prawdopodobieństwem gorzej przystosowanych.

Ewolucja zestawu podstawowych pojęć jest w nauce zjawiskiem normalnym – prowadzi do znalezienia zestawu najwygodniejszego. Czysty, ścisły opis określonego mechanizmu pozwala na wygodne dostrzeżenie zależności i wyłuskanie z nich przyczyn interesującego nas efektu. Taki model zwykle jest jednak uproszczonym opisem, pewne elementy mogą być w nim świadomie zignorowane, inne umknęły naszej uwadze. Jednak jest to jedyna droga pozwalająca uchwycić te elementy, odróżnić je od siebie nawzajem i doprowadzić model do zadowalającej, rozumiałej postaci, dającej dobrą zgodność z

doświadczeniem. Jest to szczególnie ważne i jednocześnie szczególnie trudne w biologii, gdzie badane obiekty są skrajnie złożone i posiadają bardzo wiele podobnie istotnych parametrów. Ta wielka liczba podobnie istotnych parametrów jest oczywistym utrudnieniem modelowania. Z drugiej jednak strony, jest ona warunkiem powodzenia ewolucji – trudniej wpaść w „ślepy zaułek” (lokalne optimum), który pojawia się, gdy dokładnie wszystkie parametry jednocześnie są dobre najlepiej jak można w sporym ich zakresie od osiągniętej wartości. Przy dużej liczbie parametrów staje się to po prostu mało prawdopodobne.

Złożoność sama w sobie wprowadza pewne specyficzne własności, które nie wynikają z analizy poszczególnych składowych. Jak na razie nie istnieje jedna, ogólnie przyjęta definicja złożoności. Na stronie: „Complexity_Zoo” jest obecnie 489 różnych ujęć. Wydaje się, że to cenne pojęcie ma wiele aspektów i do określonego zastosowania trzeba użyć odpowiedniego z nich. Kiedyś koncepcję redukowalności złożonego systemu do własności jego elementów nazywano redukcjonizmem, dziś badania sieci złożonych są zaawansowaną i szybko rozwijającą się dziedziną. Dziedzina ta bada właśnie owe niegdyś kwestionowane własności wynikające z parametrów sieci, które zwykle mają statystyczny charakter i opisują wzajemne powiązania – uzależnienia elementów sieci. Wiele istotnych własności żywych obiektów ma właśnie takie pochodzenie.

Wkroczyliśmy w tę nową dziedzinę na przykładzie sieci KAUFFMANA (1993) i jego modelu regulacji genowych. Jest to obecnie model bardzo popularny i intensywnie badany, choć określany jako „wczesny”, w odróżnieniu od modeli mniej abstrakcyjnych budowanych

na bazie koncepcji BANZHAF (2003). Ma on dość dobrą zgodność z doświadczeniem, jednak w tej nowej, ogromnej dziedzinie szerokość i różnorodność koncepcji i badań jest jeszcze stosunkowo niewielka, co należy brać pod uwagę przy ocenie sukcesów, które mogą być zwodnicze.

Według Kauffmana jego model potwierdza tezę, że źródłem uporządkowania obserwowanego w żywych organizmach jest oprócz mechanizmu darwinowskiego także „spontaniczne uporządkowanie” i to może nawet w większym stopniu. Koncepcja ta nosi nazwę „życie na granicy chaosu”. Jest to obecnie chyba najpoważniejszy atak na osiągnięcie Darwina. Choć nie podważa on, a nawet stosuje mechanizm darwinowski, to jednak usiłuje zaanektować znaczną część skutków (obserwowanego uporządkowania). Koncepcja ta nie ma obecnie wyraźnej opozycji, niewątpliwie jest ważna i należy do najnowocześniejszych trendów, jest bardzo szeroko i solidnie udokumentowana (co odbiera odwagę potencjalnym przeciwnikom), jednak w mojej opinii niedocenia homeostazy opartej na ujemnych sprzężeniach zwrotnych i niebawem powinna zastąpić ją jeszcze bardziej rozbudowana i głębiej zinterpretowana następna jej wersja, z powrotem doceniająca potęgę mechanizmu darwinowskiego. Moje badania (GECOW 2008a) adaptacyjnej ewolucji sieci złożonych, w tym sieci Kauffmana, wskazały źródła pewnych prawidłowości ewolucji ontogenezy. Jak zobaczymy, pojęcia: „sieć Kauffmana” i „chaos” (TEMPCZYK 1998), użyte tu przed chwilą, nie są trudne i powinny wejść do obiegu w biologii. Przyszłość wyjaśniania mechanizmów ewolucji biologicznej to właśnie ścisłe modele o charakterze statystycznym, choć złożone, tak jak złożone mogą być sieci Kauffmana.

ALGORYTMY EWOLUCYJNE

„Algorytm”, to znacznie szersze pojęcie niż „program komputerowy”, jednak zapisany w postaci programu osiąga pełną jednoznaczność. Pod pojęciem „Algorytm Ewolucyjny” powszechnie rozumie się jedynie „Algorytm Genetyczny”, który jest znacznie skuteczniejszy od innych wersji, jednak te inne, niegenetyczne algorytmy ewolucyjne istnieją i w podręcznikach są klasyfikowane. Jako narzędzie, algorytmy ewolucyjne służą przede wszystkim do optymalizacji i w podręcznikach zwraca się uwagę na ich skutecz-

ność i wygodę. My jednak rozpoczniemy od skrajnie prostego modelu, bez krzyżowania (tj. bez crossing-over), który pozwoli wyrobić sobie intuicję niezbędną do poprawnej interpretacji.

Niech ewoluujący obiekt posiada m cech, każda opisana liczbą naturalną z zakresu $[0, s-1]$, czyli mogącą przyjmować s różnych wartości. Załóżmy, że wiemy jaki wariant każdej cechy daje największe prawdopodobieństwo przeżycia. Tu Kreationista uśmiechnie się zadowolony myśląc, że całe dalsze rozumowa-

nie można już pominąć, bo wprowadzony zostaje „Inteligentny Projekt”, do którego dążyć będzie ewolucja. Nie jest jednak tak źle.

Rzeczywiście, ewolucję kierować będziemy do tego ideału za pośrednictwem warunku adaptacji, który posługując się parametrem fitness eliminuje zmiany oddalające od ideału. Jest to sposób na uproszczenie algorytmu pozwalający zastąpić symulację długiego i bardzo złożonego procesu trwania osobnika, która poprzez analizę bardzo wielu zdarzeń elementarnych ma zdecydować o jego eliminacji lub przeżyciu. Taka decyzja o eliminacji nadal ma charakter zdarzenia losowego. (Aby lepiej związać ją z cechami ewoluującego obiektu warto by uśrednić po wielu podobnych obiektach. Część z nich zginie nie wydając potomstwa, czyli zostanie wyeliminowana. To uśrednienie jest jednak obecne jedynie w naturze, gdy mutacja testowana jest przez wiele pokoleń wegetatywnych.) Sukces lub porażka wynika jedynie z przystosowania do aktualnego środowiska i fluktuacji statystycznych. Ta eliminacja jest automatyczna, działa według reguł statystycznych, nie używa jednak do tego żadnej wiedzy o „Inteligentnym Projekcie”, przeciwnie – tą drogą naprowadzania na poprawny kierunek odkrywa wiedzę o takim aktualnym ideale, który jednak zmienia się z czasem (o czym będzie dalej). Taki rzeczywisty (pełny) algorytm (proces) ewolucji nie używa parametru fitness, choć obserwując wyniki tego automatycznego (spontanicznego) procesu można go oszacować. Trzeba w tym celu uśrednić przypadki eliminacji po czasie cyklu życiowego i zbiorze podobnych obiektów (populacji). W praktyce używając algorytmów ewolucyjnych pomijamy więc modelowanie tego procesu trwania i szacowania fitness. Zastępujemy go równoważnym dla danego badanego problemu obliczeniem fitness, np. na podstawie podobieństwa do owego arbitralnie wskazanego ideału, lub np. uzyskanej wydajności jakiegoś procesu. Na podstawie tego fitness decyzja o eliminacji jest już jednoznaczna. W zastosowaniach praktycznych algorytmów ewolucyjnych zwykle poszukuje się tego ideału znając metodę określenia podobieństwa do niego. Fitness b można określić najprościej jako liczbę cech obiektu, które są idealnym wariantem tej cechy. Są oczywiście bardziej złożone definicje, ale już ta najprostsza daje pewne typowe efekty.

W algorytmie ewolucyjnym zmiany są losowe, co generuje nowe, nieprzewidziane cechy jako potencjalne kandydatki do pozo-

stania w ewoluującym obiekcie. Algorytm w naturze prowadzi ewolucję do optimum, ale „nie zna” go, choć go wskazuje i „nie wie”, dlaczego ono jest takie, a nie inne. Jak dalej zobaczymy, może być inne, ale dany konkretny proces znalazł akurat to optimum. Zbliżona w swej ścisłej naturze teoria gier ewolucyjnych (patrz artykuł ARGASIŃSKIEGO w tym zeszycie KOSMOSU) wskazuje m.in. przy użyciu dynamiki replikatorowej stabilne ewolucyjnie strategie i przyczyny tego wyboru, ale jest to wybór deterministyczny, ze z góry określonego zbioru. Do symulacji ewolucji elementów tego zbioru stosowano algorytmy ewolucyjne. Gry ewolucyjne dotyczą oddziaływania osobników w populacji (zbiórze graczy) w przeciwieństwie do algorytmu ewolucyjnego, który dotyczy ewolucji cech pojedynczego obiektu.

Powróćmy do naszego prostego przykładu. Przy pomocy fitness b definiujemy dalej warunek adaptacji (eliminacji), zazwyczaj w postaci $b_{t+1} \geq b_t$, który mówi, że fitness nie może maleć z pokolenia na pokolenie (t – numer pokolenia opisujący czas) w wyniku pojawiania się nowej zmienności, która z zasady jest losowa. Zmienność jest więc dowolna, bezkierunkowa, tzn. w każdym możliwym kierunku niezależnie od tego, co dzieć się będzie z fitness w jej wyniku. Warunek adaptacji odrzuca zaistniałe już zmiany pomniejszające fitness, tzn. eliminuje je, co odpowiada w praktyce nieudanej próbie przeżycia i pozostawienia przynajmniej jednego potomka. Jeżeli fitness – prawdopodobieństwo sukcesu – zmniejszyło się w stosunku do poziomu minimalnego wystarczającego do przeżycia, to nie powinno to nikogo dziwić. Oczywiście eliminacja jest w praktyce stochastyczna, a u nas w modelu wyznaczona jednoznacznie (można interpretować, że operujemy na wartościach oczekiwanych), ale to uproszczenie (o którym już było wyżej) niewiele zmienia a usprawnia algorytm. Eliminacja w modelowanej rzeczywistości biologicznej to śmierć, koniec. Jeżeli obiekt był tylko jeden, to proces jego ewolucji się urwał, ale jeżeli zbiór obiektów podobnych był większy (a rozmnażanie na to pozwala), to trwa on w pozostałych obiektach, aż do chwili, gdy liczebność spadnie do zera. W programie komputerowym robi się to inaczej: gdy zmiana została odrzucona, powraca się do stanu sprzed zmiany i losuje następną zmianę, aż do skutku. Jest to tzw. strategia (1+1) często stosowana w prostych symulacjach. W takim przypadku proces jest tak długi, jak życzy

sobie programista. Takiej symulacji nie pytamy jednak: jak długa okazała się ta symulacja i czy obiekt zbliża się do założonego ideału (to uwaga dla kreacjonistów, których zbyt ogólna krytyka nie zwraca uwagi na to, o co pytamy w danej konkretnej symulacji, a co jawnie zakładamy).

Zastanówmy się nad postacią warunku adaptacji, która kieruje proces do optimum, i jej interpretacyjnymi implikacjami. Stosując algorytmy ewolucyjne zwykle zakładamy, że to, do czego dopasowujemy zmienny obiekt jest stałe, a interesuje nas najlepsze dopasowanie. W ewolucji biologicznej jednak, jeżeli określone fitness b wystarcza do przeżycia i wydania przynajmniej jednego potomka, to czemu podwyższamy poprzeczkę po uzyskaniu lepszego wyniku? Rzeczywiście nie jest to konieczne, choć taki model byłby w większości przypadków nazbyt uproszczony.

Na razie powodu „podnoszenia poprzeczki” rzeczywiście nie widać, ale założenie to ma silne interpretacyjne podstawy. Prześledźmy, co dzieć się będzie z tymi obiektami, którym udało się uzyskać większe b . Z definicji uda im się pozostawić nieco więcej (statystycznie) potomstwa i szybko udział ich potomków w zbiorze wszystkich rozważanych obiektów będzie wzrastał. Rosnąć będzie także liczebność całego zbioru³. W coraz większym zbiorze prawdopodobieństwo uzyskania jeszcze większego fitness także rośnie, przypadki takie zdarzać się więc będą coraz częściej i koło się zamyka – następuje coraz szybsza eksplozja ilościowa. Zmienność zazwyczaj psuje fitness, więc obiektów blisko początkowego b będzie najwięcej, ale obiektów ze znacznie większym fitness także będzie bardzo dużo i średnie b będzie rosło. Taka sytuacja w praktyce nie może trwać długo, ogranicza ją pojemność środowiska i wynikająca z tego nieunikniona konkurencja, która podnosi wymagania stawiane obiektowi. Parametr b przestaje mieć prostą interpretację – „prawdopodobieństwo przeżycia” lub „liczba potomstwa”, raczej odpowiada mu teraz pojęcie „doskonałości”. W praktyce nadal granicą eliminacji pozostaje prawdopodobieństwo przeżycia (zawierające w sobie pozostawienie jednego potomka), ale ono systematycznie podnosi wymagania na „doskonałość” obiektu, który

ściga się z innymi w podnoszeniu tej „doskonałości”, a jest to walka o byt (życie). Pozostać może stała liczba obiektów, raczej te z największym b lub b bliskim największego, reszta musi umrzeć. I to jest interpretacyjny powód podnoszenia poprzeczki dla b , które nadal nazywać będziemy „fitness” zgodnie ze zwyczajami w omawianej dziedzinie algorytmów ewolucyjnych, w której interpretacja parametru b jest dowolna, inaczej niż w teorii ewolucji. Spadek prawdopodobieństwa przeżycia (pierwotnego określenia fitness) obiektu, który nie zmienił się (i ma ciągle takie samo b zależne tylko od niego) wynika ze zmiany środowiska powodowanej konkurencją. Parametr b opisuje jedynie stan obiektu, jego doskonałość, a o przeżyciu decyduje w warunkach konkurencji także doskonałość innych obiektów. Konkurencja o deficytowe dobro może podsuwać rozmaite rozwiązania, jednym z nich jest zastąpienie go przez inne, na razie nie wykorzystywane dobra, co na pewien czas likwiduje konkurencję, pozwala na chwilową eksplozję ilościową. Jest to jedna z dróg wzrostu doskonałości b . Obraz ten (zmiana interpretacji b i rozpad jej na dwie wielkości z ww. ich okolicznościami) pozwala uświadomić sobie elementy problemu „postępu ewolucyjnego” (GECOW 2008b). Powyższy opis zakładał stały ideał przystosowawczy, co jednak w warunkach zmiennego, jak widzimy, środowiska nie jest słuszne w dłuższym przedziale czasu. Środowisko w algorytmie ewolucyjnym występuje jedynie w formie warunku eliminacji.

Ten skrajnie prosty algorytm ewolucyjny zwykle stosowany jest do badań o charakterze rozpoznawczym i jakościowym. Stosowałem go do ww. symulacyjnych badań prawidłowości ewolucji ontogenezy (GECOW 2008a). Stosował go też BANZHAF (2003) – ojciec omawianych na końcu biologicznych modeli regulacji genowych. Ze względu na niską wydajność algorytm taki nie ma jednak większego praktycznego znaczenia dla celów rzeczywistej optymalizacji. W rzeczywistości biologicznej dotyczy organizmów rozmnażających się tylko wegetatywnie, bez żadnych mechanizmów „poziomej” (między współistniejącymi osobnikami) wymiany informacji genetycznej.

³Biolodzy nazywają także ten zbiór „populacją”, ale to całkiem inny zbiór i związany z nim mechanizm, niż w przypadku obecności krzyżowania i używanie tej samej nazwy jest bardzo mylące. Ja więc tego nie popieram i jak mogę – nie czynię. Jest, co prawda, nazwa: „populacja mendlowska” dla odróżnienia tej drugiej, ale ta nazwa jest za długa więc nie jest (i nie będzie) praktycznie używana.

W zastosowaniach metod optymalizacyjnych, do których należą algorytmy ewolucyjne, w pierwszej kolejności trzeba sformułować „fitness”, inaczej funkcję celu, tj. wartość ocenianego parametru w postaci funkcji użytych tu losowo zmiennych „cech” jako parametrów. Powyżej była to liczba cech zgodnych z ideałem. Optymalizacja to znajdowanie położenia ekstremum, (inaczej: optimum) – maksimum jak w fitness o biologicznej interpretacji lub minimum jako odległość od zadanego ideału. Zwykle w zadaniach optymalizacyjnych jest to minimalizacja. Owo ekstremum funkcji celu szukane jest w przestrzeni parametrów poprzez przypadkowe błędzenie i odrzucanie ruchów o niewłaściwym skutku. Wartości takiego „fitness” (niezależnie od wyżej omawianego kierunku) tworzą zwykle bardzo złożony krajobraz w przestrzeni parametrów,

posiadający wiele optimumów o różnej wartości ekstremalnej. W przypadku biologicznego fitness nazywa się je krajobrazami Wrighta. Proste losowe błędzenie pozwala znaleźć jedno takie lokalne optimum, po czym proces zatrzymuje się; nie wiadomo jednak, czy jest to optimum najlepsze (globalne), nawet nie wiadomo, czy należy do lepszych. Problem „przeskoczenia” do innych optimumów jest trudny i istnieje wiele różnych metod jego częściowego pokonania. Podobne problemy ma rzeczywista ewolucja. Z powodu ciągłej konkurencji (wewnątrzgatunkowej i pomiędzy gatunkami) minimalne fitness b wystarczające do przeżycia podobne jest do powierzchni stale podnoszącej się wody. Szczepy, które znalazły się na takiej górze⁴ bez wyjścia czeka utopienie.

ALGORYTMY GENETYCZNE

Podstawową różnicą i wyższością algorytmu genetycznego nad resztą algorytmów ewolucyjnych jest zastosowanie w nim krzyżowania, pozwalającego na składanie osiągnięć z równoległych procesów w jednym osobniku. Algorytm ten odtwarza dość dokładnie rzeczywisty mechanizm ewolucji biologicznej (GOLDBERG 1995) – można powiedzieć, że u organizmów rozmnażających się płciowo ewolucja stosuje ten algorytm. Jest on określony dość ogólnie, jest bardzo wiele konkretnych realizacji rozmaicie definiujących pozostałe szczegóły i tu zastosowania numeryczne różnią się znacznie od tego, co spotykamy w naturze; zwykle są to silne uproszczenia. Algorytm stosowany przez naturę różni się od numerycznego przede wszystkim: statystycznie rozmytą eliminacją – zarówno w odniesieniu do progu fitness jak i czasu, co związane jest ze statystycznym charakterem wymagań środowiska i jego niejednorodnością; populacjami o niedokładnym odgraniczeniu ze względu na migracje. Różni się też ogromną różnorodnością form zmienności. Podstawowy algorytm genetyczny został zdefiniowany przez Johna Hollanda z Uniwersytetu w Michigan i rozwinięty dzięki jego pracom z lat 60 i 70.

W Algorytmie Genetycznym mamy populację osobników. Osobnik ma geny – konkretne liczbowe parametry opisujące jego cechy, ułożone kolejno na chromosomie (wektorze genów). Na podstawie tych genów (tj. parametrów) obliczane jest fitness (funkcja celu).

Odrzuca się określoną część osobników o najmniejszym fitness, a na ich miejsce wprowadza osobniki potomne. Powstają one z wylosowanej pary rodziców, poprzez mutacje i krzyżowanie. Zwykle preferuje się rodziców z większym fitness a parametry mutacji (liczba, miejsce, zakres zmiany wartości) oraz rekombinacji (punkt(y) na chromosomie) są losowane. Po wyprodukowaniu nowych osobników potomnych (właściwie ich chromosomów) na miejsce wcześniej odrzuconych, ponownie obliczane jest fitness i znowu odrzucana jest określona część osobników o najmniejszym fitness. Jest to strategia $(\mu+\lambda)$, gdzie ogólnie populacja zawiera μ osobników (po redukcji, potencjalnie dopuszczonych do rozrodu), a w wyniku krzyżowania powstaje λ nowych (potomnych) osobników. Należy przy tym uświadomić sobie, że ten ogólny opis ma wiele niezdefiniowanych szczegółów, które mogą się różnić w konkretnych rozwiązaniach.

Klarowne tu kroki algorytmu są rozmyte w rzeczywistości biologicznej, decyzje oparte o wartość fitness są w algorytmie pewne a w naturze też statystycznie rozmyte, ale właśnie ta jakby nienaturalna klarowność algorytmu pozwala zrozumieć zasadę i zobaczyć jej skutki.

Równoległość „obliczania” wielu osobników (zawierającego kolejno ich zmienność i testowanie) występująca w rzeczywistej populacji zamieniona zostaje w typowym komputerze na sekwencyjność obliczeń i ta

⁴Optimum biologicznego fitness to jego maksimum, a nie minimum jak w zastosowaniach technicznych.

potencjalna równoległość ogólnego algorytmu nie daje zysku na czasie obliczeń numerycznych, ale i tak zazwyczaj algorytm ten obliczany w komputerze okazuje się szybszy od podobnego bez krzyżowania. Oznacza to ogromną przewagę w biologicznej rzeczywistości, gdzie zysk tempa ewolucji na tej równoległości występuje. Uważam, nie jest to jednak pogląd nowoczesny, że ten zysk tempa jest jedną z głównych przyczyn dlaczego kosztowne mechanizmy rozmnażania płciowego są tak powszechne a „pomysły” niektórych gatunków by te koszty zredukować poprzez partenogenezę – prawie zawsze tak niedawne ewolucyjnie. Szerzej o tym złożonym i ciągle dyskusyjnym zagadnieniu pisze KORONA (1998).

W algorytmie genetycznym właścicielem („magazynem”) cech, często alternatywnych w kilku wykluczających się wariantach (allele), jest populacja a nie osobnik i wystarczająca reakcja (polegająca na powstaniu osobników o dostatecznym fitness) na nagłe zmiany wymagań środowiska może być znacznie szybsza (wystarczy często inaczej zestawić już istniejące w populacji elementy) niż w przypadku rozmnażania wegetatywnego, gdzie te wszystkie rozwiązania muszą w tym czasie zostać znalezione przez zmienność mutacyjną w pojedynczej nitce ewolucyjnej (co prawda – jednej z wielu).

W algorytmie genetycznym też występuje problem optimów lokalnych, na który znaleziono wiele sposobów, jednak żaden nie rozwiązuje problemu do końca. Ewolucja biologiczna także często utyka na lokalnych optimach, co w warunkach ciągłej konkurencji (także międzygatunkowej) prowadzi do wymarcia populacji. Jedną z podstawowych nadziei takich „powodzian” (patrz koniec poprzedniego rozdziału) zgromadzonych na szczycie górkę są „ruchy górotwórcze” – krajobraz Wrighta, w którym istnieją owe górkę fitness jest dość zmienny, czuły na zmiany w środowisku nie tylko abiotycznym, ale przede wszystkim na zmiany w środowisku biotycznym i taka droga ratunku jest dość prawdopodobna. (W którymś z wielu wymiarów tej przestrzeni górkę zanika i otwiera tym nową ścieżkę pod górę.) Proszę zauważyć, że nowe

rozwiązanie, np. zastąpienie deficytowego substratu innym, także zmienia ten krajobraz i to nie tylko dla „wynałazcy” – na tym samym obszarze będzie teraz więcej osobników i konkurencja przesunie się na inne dobra.

Skuteczność algorytmów genetycznych, jako narzędzia optymalizacji, jest niewątpliwie potwierdzeniem potęgi doboru naturalnego wskazanego przez Darwina. Kreationiści uważają jednak to twierdzenie za błędne – podobnie w odniesieniu ogólnie do algorytmów ewolucyjnych uważają, że aby algorytm taki rozwiązał problem, trzeba inteligentnie go zadać, głównie odpowiednio dobrać kryteria oceny, wstawiając tym niejako docelowy wynik w konstrukcję algorytmu. Skrajna odpowiedź na ten zarzut prowadzi do stanu: „strzyżone-golone”. Kryteria oceny w ewolucji biologicznej, tak jak i w poprzednim rozdziale, są proste i automatyczne – zawsze tylko „przeżycie” (w sensie pozostawienia po sobie potomstwa). Obserwujemy i podziwiamy tylko to, co dotrwało. Owszem, w konstrukcję algorytmu doboru naturalnego jest włożony ten docelowy wynik – istnienie, ale nic poza tym, żaden konkretny „Inteligentny Projekt” dokładnie opisujący konstrukcję jakiegoś zwierzęcia lub jego części. Pokazaliśmy ponadto, że ten projekt gdyby istniał, musiałby się sukcesywnie zmieniać wraz ze zmianami środowiska. Tak podziwiana budowa oka jest wynikiem działania algorytmu, nie jest w niego wbudowana. Pozostaje „inteligentna” konstrukcja samego algorytmu i realizujących go składników. Tego Darwin nie wskazał. Nie musiał, wystarczy, że wskazał dobór naturalny, a pochodzenie tego mechanizmu to problem osobny, problem powstania życia. Wskazany dobór naturalny to jednak ogromna, zasadnicza redukcja problemu pochodzenia obserwowanej złożoności i celowości budowy obiektów żywych. Szanse na spontaniczne powstanie mechanizmów darwinowskich są znaczne, gdyż sam algorytm nie jest złożony, wydaje się nawet, że wymiana informacji tworząca silniejszy z tych mechanizmów, była obecna już od początku, kiedy powstawały pierwsze hipercykle (patrz artykuł WEINERA w tym zeszycie KOSMOSU).

SIECI KAUFFMANA

Sieć i graf to w zasadzie to samo. Gdy mówimy o grafie to używamy pojęć „wierzchołek” i „krawędź”, która łączy wierzchołki;

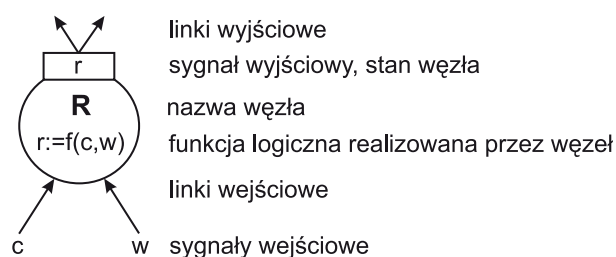
gdy mówimy o sieci, na wierzchołek mówimy „węzeł” a na połączenie: „łuk”, „link” itp. Oba ujęcia typowo są przeplatane w jednej

wypowiedzi, więc nie będziemy i my zwracać uwagi na konsekwencję w tym aspekcie. Sieci skierowane charakteryzują się przypisaniem kierunku (dokładniej – zwrotu) do połączeń, co na rysunkach zaznacza się strzałką. Sieć Kauffmana jest siecią skierowaną.

KAUFFMAN (1969) skonstruował swoją sieć do celów opisu zjawisk biologicznych, jednak dostrzegli jej walory dopiero fizycy i to po wielu latach. Sieć Kauffmana ma ważną dodatkową cechę – funkcjonuje (mówi się: ma dynamikę), tzn. węzły tej sieci mają wejścia, na które podawane są sygnały wejściowe i wyjścia, na których pojawia się sygnał wyjściowy, tzw. stan węzła, jako wynik funkcji, której argumentami są sygnały wejściowe (Ryc. 1). W zasadzie do badań używane są tylko sygnały o dwóch wariantach, najczęściej opisywanych jako 0 i 1 i interpretowanych jako stany logiczne (KAUFFMAN 1993), dlatego sieci Kauffmana są niemal synonimem sieci logicznych (Boolowskich).

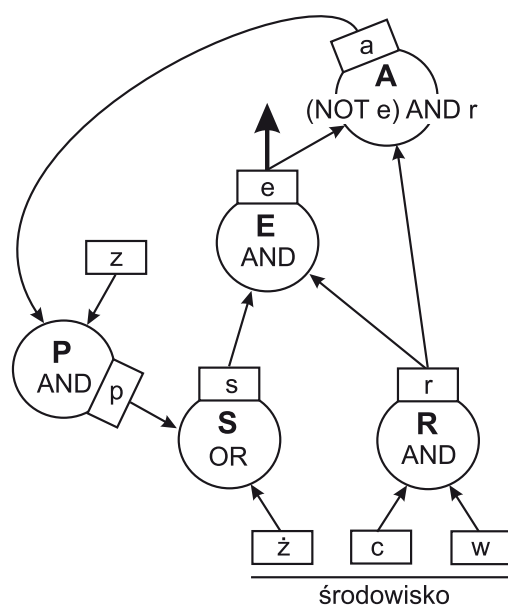
Typowo sieć Kauffmana obliczana jest synchronicznie, co oznacza, że w jednym kroku czasowym obliczane są nowe stany wszystkich węzłów, używając jako argumentów funkcji sygnałów wejściowych wygenerowanych w poprzednim kroku czasowym, tj. stanów węzłów, które są połączone linkiem do danego wejścia. W takim sposobie kolejność liczenia węzłów nie ma znaczenia, a wynik jest jednoznaczny, tzn. sieć jest deterministyczna (patrz Ryc. 2 i Tabela 1).

Zazwyczaj rozważa się sieci autonomiczne, tzn. nie posiadające wolnych wejść i wyjść, czyli bez kontaktu z otoczeniem. Daje to wygodną sytuację teoretyczną. Sieci są skończone – mają skończoną liczbę węzłów.



Ryc. 1. Węzeł sieci w sieci Kauffmana i jego elementy.

Przekształca on sygnały wejściowe podawane przez linki wejściowe w chwili poprzedniej, używając funkcji logicznej. Wynik funkcji zastępuje stan węzła z chwili poprzedniej i stanowi sygnał wyjściowy, podawany w obecnej chwili poprzez linki wyjściowe do ich odbiorców, gdzie stanowią sygnały wejściowe z obecnej chwili.



Ryc. 2. Przykładowa (wymyślona) sieć regulacji genowej.

Sygnały (opisane w Tabeli 1) przyjmują jedynie dwa stany: jest = 1 oraz brak = 0. Odpowiada to aktywności genów reprezentowanych przez węzły (R, E, A, P). Węzeł S wynika z reguły opisu – łączy sygnały p i z w jeden sygnał wejściowy s do węzła E. Tabela 1 pokazuje funkcjonowanie sieci w kolejnych chwilach czasu (krokach czasu).

Stanem sieci jest stan wszystkich jej węzłów. Po każdym kroku czasowym stan sieci jest inny, ale wszystkich stanów jest skończona liczba, więc sieć musi kiedyś osiągnąć stan, w którym już była i odtąd cyklicznie będzie go osiągać. Ciąg stanów sieci (trajektoria) po osiągnięciu pierwszego powtarzającego się stanu jest więc zamkniętą linią w przestrzeni stanów sieci – nazywamy go atraktorem cyklicznym. Nie każdy stan leży na atraktorze – zbiór stanów, z których osiągnany jest dany traktor, nazywany jest basenem atrakcji (przyciągania).

Kauffman uważa, że zróżnicowanie komórek tkanek jest wynikiem ustabilizowania się różnych atraktorów w sieci interakcji międzygenowych jednokowego genomu. Bada więc w zależności od parametrów sieci liczbę i długość atraktorów oraz łatwość powrotu do tego samego atraktora po małym zaburzeniu. Głównymi parametrami są tu liczba K – wejść do węzła, którą zwykle ustalał dla całej sieci (obecnie bada się zmienne K) oraz N – liczba węzłów w sieci, która w pierwszym przybliżeniu odzwierciedla złożoność sieci. Rozkład liczby k – wyjść z wę-

Tabela 1. Funkcjonowanie sieci z Ryciny 2 w pełni wyznacza sieć i stan z chwili 0 wraz z zadanymi stanami środowiska i wielkości zapasu.

W zimie komórka (powiedzmy – glonu) miała jakieś zapasy i leżała na podłożu z żywnością, ale nie rosła (chwila 0). Gdy zegar długości dnia wskazał wiosnę, było jeszcze za zimno (chwila 1). Zrobiło się dostatecznie ciepło (chwila 2) i gen R stał się aktywny – odtąd produkuje aktywator wzrostu r. Gen E ma od dawna substrat z pożywienia \hat{z} ale dopiero teraz (chwila 3) dotarł aktywator r więc w następnej chwili 4 występuje efekt e w postaci wzrostu. Chwilę wcześniej (chwila 3) zabrakło substratu w pożywieniu \hat{z} na zewnątrz, ale wewnątrz jeszcze był substrat s, który został wykorzystany do efektu w chwili 4. Jednak nowego s w chwili 3 już nie dostarczono do środka, a zatem w chwili 4 już go brak, więc w następnej chwili 5 nie ma już efektu e, co dopiero w chwili 6 poprzez gen A wszczyna alarm a, który aktywuje produkcję P z zapasów z. Produkt p pojawia się więc w następnej chwili 7 i dociera poprzez S do E w chwili 8, co daje efekt w chwili 9, a ten wyłącza w chwili 10 alarm, po czym ustaje produkcja w chwili 11. Brak substratu w chwili 12 poprzez brak efektu w chwili 13 włącza znów alarm w chwili 14, następnie produkcję w chwili 15, co wyczerpuje zapasy (chwila 15), ale to nie jest tu opisane siecią, i daje efekt (chwila 17). Brak zapasów powoduje dalszy ciąg zdarzeń zaznaczony w tabeli pogrubieniem.

chwila	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
sygnał																							
ciepło	c	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
wiosna	w	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
żywność	\hat{z}	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
zapas	z	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
substrat	s	1	1	1	1	0	0	0	0	1	1	1	1	0	0	0	0	1	0	0	0	0	0
rosnąć	r	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
efekt	e	0	0	0	0	1	0	0	0	0	1	1	1	0	0	0	0	1	0	0	0	0	0
alarm	a	0	0	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1	0	1	1	1
produkt	p	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	1	0	0	0	0	0	0

złów (zwanej stopniem wierzchołka) charakteryzuje typ sieci. Do końca ubiegłego wieku badano sieci Erdősa-Rényi, zwane „Random” (ang. Random Boolean Network, RBN), o dzwonowatym rozkładzie stopnia wierzchołka (ALBERT i BARABÁSI 2002). Obecnie modne są sieci bezskalowe Barabásiego-Alberta (BARABÁSI i BONABEAU 2003) (SFRBN – patrz IGUCHI i współaut. 2007), dla których rozkład stopnia wierzchołka na wykresach log-log jest prostą. Powstają one poprzez preferencyjne dodawanie – nowy węzeł dołącza się do węzłów obecnych już w sieci z prawdopodobieństwem proporcjonalnym do ich „stopnia”, tj. liczby linków. Okazuje się, że większość zbadanych rzeczywistych sieci ma taki rozkład stopnia wierzchołka. Badane są także inne sieci, w tym o więcej niż dwóch wariantach sygnału. Istotnym warunkiem prezentowanych poniżej rozważań ilościowych jest poprawna losowość funkcji przypisanych węzłom oraz założenie jednakowego prawdopodobieństwa obu wariantów sygnału. Przedstawianie wyników otrzymanych bez tych

założeń oraz dla innych sieci niż typu RBN wykracza poza ramy tego artykułu.

Na statystyczną długość i liczbę atraktorów znalezione zostały dokładne wyrażenia zależne od K i N na drodze teoretycznej. Uzupełniają je w najważniejszym zestawieniu cech sieci w zależności od K efekty małego zaburzenia – po zmianie jednego stanu jednego elementu w sieci:

- stabilność homeostatyczna – tendencja do powrotu do tego samego atraktora. Uwaga: nie polega ona na szczególnym nagromadzeniu ujemnych (regulacyjnych) sprzężeń zwrotnych charakterystycznych dla organizmów żywych (KORZENIEWSKI 2001, 2005);
- zasięg zmian atraktora – liczba innych atraktorów, do których przeskakuje system po wszystkich możliwych takich małych zmianach;
- typ zachowania sieci, który może być: chaotyczny – gdy typowym efektem jest lawina zmian; uporządkowany – gdy lawiny nie występują, a zmiany są małe lub ich brak; przejściowy – lawiny bywają, ale są małe.

Analizując otrzymane wyniki Kauffman uważa przede wszystkim, że ograniczenie się systemów do bardzo małej liczby atraktorów o bardzo małych długościach jest wielkim ograniczeniem czyli spontanycznym uporządkowaniem. Dla porównania: przestrzeń stanów posiada ogólnie 2^N elementów, natomiast gdy $K=N$ dla $N=200$ jest tylko około 74 atraktorów a dla $N=10000$ tylko 3700. Dla $K=2$ to ograniczenie jest istotnie większe, gdyż dla $N=10000$ spodziewamy się jedynie około 100 atraktorów, a system spontanicznie wpada w jakiś atraktor o długości, która wynosi także tylko około 100; zatem z 2^{10000} stanów ogranicza się do zaledwie 100. System o $N=100000$ porównywalny z ludzkim genomem ma około 317 atraktorów. Jeżeli obserwować będziemy dwa systemy różniące się kilkoma stanami, to zwykle różnica ta będzie rosła z czasem, aż osiągnie pewien określony poziom nieco poniżej połowy N , ale nie w przypadku $K=2$, w którym nawet znaczna początkowa różnica z czasem zanika!

Dla $K=2$ stabilność homeostatyczna atraktorów wynosi około 80 do 90%, a gdy system już powróci do tego samego atraktora (stabilność homeostatyczna), to zazwyczaj znajdzie się w miejscu tego atraktora, w którym byłby bez tego zaburzenia. Większość przypadków, w których stan systemu zmienia jednak basen atrakcji, prowadzi do tego samego nowego atraktora. Większość elementów, zwykle ponad 70%, wpada w stan stały (stała wartość sygnału wyjściowego węzła), który jest jednakowy dla większości atraktorów. Są to tzw. obszary zamrożone, które tworzą spójny obszar (perkolują) pozostawiając niezależne jeziora elementów zmieniających się. Dla $K>2$ jest odwrotnie – obszar spójny zawiera elementy zmieniające się, a w nim pozostają wyspy elementów zamrożonych. Ten obraz tłumaczy wielkość zmiany (ang. damage) jaka powstaje po małym zaburzeniu – lawina zmian rozwija się bez przeszkód gdy $K>2$ a dla $K=2$ napotyka na izolujące ściany „łodu”. Różnice pomiędzy $K=2$ a $K>2$ to przejście fazowe między porządkiem a chaosem – radykalna zmiana zachowania się systemów. Dokładne wyznaczenie punktu tego przejścia fazowego dla różnych sieci (różniących się rozkładem k – wyjściowego stopnia wierzchołków przy obecnie także zmiennym K – stopniu wejściowym) jest obecnie głównym badanym zagadnieniem.

Rozważania prowadzone na nieco prostszym modelu zależności genowych zwanym NK dotyczą własności krajobrazu fitness.

Mamy tu N genów występujących w A wariantach, każdy zależy od K spośród pozostałych. Fitness genomu jest sumą fitness każdego genu z uwzględnieniem fitness tych genów, od których ten gen zależy. Dla $K=0$, czyli niezależnych genów, krajobraz ten zawiera jedną gładką górę – z każdego genotypu jest ścieżka poprzez pojedyncze mutacje na sam szczyt, a sąsiedzi (osiągnięci przez pojedynczą mutację) mają fitness bardzo zbliżone. To ostatnie powoduje, że znając fitness danego genotypu dużo wiemy o sąsiadach, ich fitness nie jest dowolne, czyli jest silnie skorelowane. Dla drugiej skrajności $K=N-1$ krajobraz jest w pełni przypadkowy (nieskorelowany) – nic nie można powiedzieć o sąsiadach (brak korelacji), jest szczególnie wiele lokalnych optimum fitness, $2^N/(N+1)$ dla $A=2$, a ogólnie $A^N/(N(A-1)+1)$, droga do nich jest bardzo krótka, a oczekiwana część dostępnych mutacji jednogennych jaką stanowią mutacje adaptacyjne zmniejsza się o połowę po każdym adaptacyjnym kroku. Genotypy mają mały wybór wśród lokalnych optimum, a do danego optimum może dojść mało genotypów. Jest to krajobraz poszarpany, surowy.

Własności krajobrazu fitness zmieniają się stopniowo, gdy K wzrasta od $K=0$ do $N-1$. Dla $K=2$, $A=2$ i większych N , lokalnych optimum jest już sporo, jednak leżą wszystkie blisko siebie i tworzą jakby „Masyw Centralny”. Ze wzrostem N , które jest najprostszą miarą złożoności, wysokość optimum spada, co wynika ze sposobu obliczania fitness i udziału genu w globalnym fitness – zmniejszają się różnice fitness pomiędzy genotypami. Prowadzi to tzw. katastrofy złożoności, tj. do spadku wpływu selekcji darwinowskiej na położenie systemu w coraz bardziej wyrównanym krajobrazie fitness. Kauffman uważa, że znaczenia nabiera wtedy samoorganizacja. Etapy optymalnej ewolucji adaptacyjnej przewiduje on następująco. Na początku, gdy fitness jest jeszcze małe, mimo skorelowanego krajobrazu długie skoki dają szybki wzrost fitness, ale czas oczekiwania na następny sukces za każdym razem podwaja się i szybko staje się on bardzo długi. Następuje drugi etap, w którym do głosu dochodzą małe zmiany wykorzystujące korelację krajobrazu. Prowadzą one do lokalnego optimum, gdzie proces zatrzymuje się. Wtedy znowu pojawia się długi skok, po którym znowu następuje dochodzenie do lokalne-

go optimum i oczekiwanie w nim na następny duży skok. Uważa on, że taki obraz odpowiada w przypadku filogenezy etapom radiacji i stasis. Długie skoki bez spadku fitness są jednak dla większych fitness wyjątkowo mało prawdopodobne i należy je traktować jako jedyny ratunek (który jednak rzadko się zdarza) po zatrzymaniu się w lokalnym optimum. Są one bardzo widoczne, ale główny wzrost fitness odbywa się małymi zmianami.

Własnościami krajobrazu są przede wszystkim: liczba lokalnych optimum i ich wzajemne usytuowanie, co opisuje stopień korelacji. Duże korelacje występują dla małych K , a dla dużych K krajobraz jest nieskorelowany (surowy), co daje chaotyczne zachowanie systemu nawet po mutacji na jednej pozycji. Struktura krajobrazu, gładka lub surowa, rządzi zarówno „*evolvability*” populacji, jak i utrzymywaniem fitness jej członków, a także stawia ograniczenia poszukiwań adaptacyjnych. Dla gładkiego krajobrazu, populacji o stałym rozmiarze i stałej częstości mutacji, gdy wzrasta złożoność (N) osiągnąć jest próg błędów i populacja „rozplywa się” poprzez prawie neutralne mutacje na większym obszarze wokół piku. W krajobrazie dostatecznie surowym poszukiwania populacji ograniczają się do bardzo małego obszaru. Gdy wzrasta częstość mutacji, pojawia się kilka „przebiegów fazowych”, po których populacja także rozlewa się na większy obszar. Możliwość optymalizacji „*evolvability*” pozwoliłaby populacji znacznie zwiększyć zakres poszukiwań.

Krajobraz fitness nie jest stały, lecz zmienia się w wyniku oddziaływań pomiędzy współewoluującymi obiektami. Daje to możliwość uwolnienia z pułapki lokalnego optimum. Hierarchia wśród oddziałujących genów pozwala organizmowi dostroić swoją odpowiedź do stopnia deformacji krajobrazu fitness. Od charakteru krajobrazu zależy także użyteczność krzyżowania.

Konkretne sieci Kauffmana (systemy) posiadają różne parametry, zazwyczaj są one ustalone. Należą tu oprócz K i N także opis połączeń i przypisanie funkcji wierzchołkom. Podanie ich kompletu dokładnie definiuje system – jest to punkt w prze-

strzeni parametrów systemu. Punktem temu odpowiada konkretny zbiór basenów atrakcji i atraktorów. Zazwyczaj poruszając się nieznacznie w przestrzeni parametrów uzyskuje się nieznaczne zmiany własności systemu, ale są granice, których przekroczenie powoduje radykalne zmiany, np. znika lub pojawia się jakiś basen atrakcji. Takie granice zwane są bifurkacjami, tworzą powierzchnie dzielące przestrzeń parametrów na zwykle spore obszary systemów o określonym zachowaniu dynamicznym zwanych „*strukturalnie stabilnymi*” (THOM 1970, KAUFFMAN 1993).

Można wyobrazić sobie, że w pewnych aspektach ewolucja adaptacyjna to błądzenie (ang. *adaptive walks*) po przestrzeni parametrów w poszukiwaniu „dobrego” zachowania dynamicznego. Systemy strukturalnie stabilne ewoluują w krajobrazie skorelowanym, a chaotyczne, które nie są stabilne strukturalnie – w nieskorelowanym krajobrazie, gdzie mała zmiana powoduje zwykle przejście przez wiele bifurkacji zmieniając radykalnie zachowanie systemu. Szansa na sukces takiej zmiany jest znikoma, a dla większych fitness – pomijalna. Ogólnie Kauffman dochodzi do wniosku, że ewolucja adaptacyjna jest skuteczniejsza w krajobrazie skorelowanym, więc poszukiwania „dobrego” zachowania dynamicznego doprowadzą do brzegu chaosu i porządku, w okolice $K=2$, gdzie stabilność strukturalna jest największa. To właśnie mówi znana hipoteza Kauffmana: „*życie na granicy chaosu*”. Jednocześnie w tym obszarze, jak było omawiane, najsilniejsze jest spontaniczne uporządkowanie. Ta stabilność strukturalna badana była wyżej poprzez małe zmiany stanów (co daje podobne wyniki jak zmiany struktury) – czy prowadzą one do wytrącenia z początkowego basenu atrakcji, czy też system powraca do uprzedniego traktora, nawet bez zmiany jego fazy. Nie można odmówić słuszności wnioskowi, że ewolucja adaptacyjna powinna dążyć do tego obszaru, można jednak powątpiewać w realne możliwości wyboru i konstrukcji kryterium eliminacji w tym zakresie oraz statystyczną obojętność pozostałych okoliczności.

MODELE REGULACJI GENOWEJ

Model Kauffmana regulacji genowych oparty na sieci Kauffmana (tj. boolowskiej czyli logicznej) (KAUFFMAN 1969, 1971, 1993),

powstał formalnie w 1971r., ale praktycznie za jego początek należy uznać 1969 r. Obecnie określany jest on jako „wczesny”. Wykazał

on możliwość opisu, symulacji i badania teoretycznego sieci uzależnień pomiędzy genami funkcjonującej równolegle (jednocześnie) i dającej zróżnicowane zachowania cykliczne przy bardzo prostych założeniach. Dwa stany logiczne występujące w tej sieci interpretowane są jako stany: aktywny lub nieaktywny danego genu – węzła sieci. Rycina 2 i Tabela 1 są poglądowym (wymyślonym, uproszczonym) przykładem takiego opisu regulacji genowej. Model ten jest głównym „doświadczalnym” potwierdzeniem hipotezy „życie na granicy chaosu”. Parametr K wyznaczony jest na podstawie porównania stabilności po małym zaburzeniu (WAGNER 2001; SERRA i współaut. 2004, 2007; RÄMÖ i współaut. 2006). Okazuje się, że przewidywania stabilności teoretyczne i symulacyjne dla $K=2$ dobrze zgadzają się z eksperymentem, w którym w rzeczywistym genomie *Saccaromices cerevisiae*, zawierającym $N=6312$ genów, wyłączano po jednym 227 genów (HUGHES i współaut. 2000). Obserwowaną tu stabilność interpretuje się jako „stabilność homeostatyczną”, nie uwzględniającą szczególnego nagromadzenia ujemnych sprzężeń zwrotnych (aktywnej kontroli dopuszczalnego zakresu zmienności parametru) charakterystycznego dla organizmów żywych. To nieuwzględnione szczególne nagromadzenie jest tak charakterystyczne, że zostało nawet użyte do jednej z wersji definicji życia (KORZENIEWSKI 2001, 2005) i pominięcie go w modelu prowadzi do przyjęcia niewłaściwego źródła stabilności (GECOW 2009). (Dla przypomnienia – ujemne sprzężenia zwrotne realizuje np. termostat w lodówce pilnując by temperatura mieściła się w zadanym zakresie.)

Obecnie uważa się ten kauffmanowski model regulacji genowych za pierwsze zgrubne przybliżenie, szczególnie po próbach skonstruowania konkretnych rozpoznanych sieci. Główne kierunki jego modyfikacji to wprowadzenie więcej niż dwa warianty sygnału, opisujące różne stężenie produktu genu, co m.in. realizuje sieć RWN (ang. Random Walk Network) (LUQUE i współaut. 2004).

Przykład sieci Kauffmana dobrze obrazuje, czym zajmuje się „teoria sieci złożonych”, w tym bezpośredni związek dużej części tych rozważań z ewolucją biologiczną. Obecnie nieproporcjonalnie szybko rozwija się część teoretyczna tych badań, za nią z trudem nadąża symulacja, chwilami tylko wyprzedzając teorię. Natomiast interpretacja, szczególnie biologiczna i ewolucyjna, ma wyraźne braki głównie z powodu omijania tej tematyki

przez biologów zajmujących się ewolucją. Samoorganizacja tak eksponowana przez Kauffmana i próbująca odebrać część terytorium skutkom mechanizmów darwinowskich ma w mojej ocenie słabe podstawy interpretacyjne. Podobnie teza „życie na granicy chaosu” może i sprawdza się w przypadku logicznych sieci regulacji genowych, ale trudno mi sobie wyobrazić, że słuszna jest bardziej ogólnie i osobiście lokują organizmy żywe w obszarze dojrzałego chaosu. Występująca tam lawina zmian w funkcjonowaniu sieci uruchamiana małą zmianą (np. struktury sieci) dobrze modeluje śmierć organizmu po mutacji niezbędnej dla darwinowskiej eliminacji.

Ośmieleni wynikami kauffmanowskiego modelu regulacji genowych genetycy zbudowali następną generację modeli (ang. Genetic Regulatory Network, GRN), które określane są jako „biologiczne”. Są one znacznie mniej abstrakcyjne i bardziej konkretne – ich elementy bezpośrednio odpowiadają realiom genetyki. Większość z tych modeli opiera się na pracy BANZHAF (2003). Modelowana jest komórka z genomem i białkami. Model zawiera np. stałą, niewielką liczbę genów. Gen opisuje kilka białek i ma kilka miejsc sterowania (cis-moduły), do których dopasowuje się odpowiedni inhibitor lub aktywator. Każde białko może wystąpić w takiej roli (być sterujące). Gen to ciąg liczb lub tylko bitów, posiadający kierunek podobny do kierunku od 5' do 3' w DNA. Białka posiadają jedynie informację o specyficzności jako białko sterujące do określonego genu. Ich głównym parametrem jest liczba ich cząsteczek odpowiadająca stężeniu. Podczas kroku czasowego ustalają się rekurencyjnie nowe stężenia białek przez uwzględnienie rozpadu i dodanie nowych białek wyprodukowanych w danym kroku czasowym (zgodnie z aktualnym stanem aktywności genów). Środowisko może oddziaływać na stężenie określonych białek, a stężenia pewnych innych białek są odczytywane jako oceniany sygnał wyjściowy.

Taki opis jest zupełny – pozwala przewidywać dalszy przebieg procesu zmian stężeń białek – tzw. dynamikę. Nie ma w nim opisu transkrypcji, intronów, RNA ani translacji zmieniającej alfabet zapisu. Te elementy są zbędne do badania dynamiki, a także ewolucji takiej sieci. Nie są tu istotne szczegóły, które różnią się w różnych modelach tego typu, a jedynie koncepcja i typ zadań, które zasadniczo odróżniają to podejście od sieci i badań Kauffmana. Nie widać i nie rozważa się w tym podejściu zagrożeń dla obszaru wy-

jaśnianego przez mechanizmy darwinowskie, a wręcz przeciwnie – modele i symulacje korzystają z nich jako z podstawowego założenia. Jako mechanizm ewolucyjny stosowane są oba wyżej opisane Algorytmy Ewolucyjne: prostszy (strategia (1+1)) – ten stosował Banzhaf w swej ogólnej pracy oraz Algorytm Genetyczny (strategia ($\mu+\lambda$)) zastosowany w poniższym konkretnym przykładzie.

W eksperymencie symulacyjnym, badającym powstawanie wewnętrznego zegara (KNABE 2006), jest 5 lub 9 genów, gen startuje opisując jedno białko i posiadając jedno miejsce sterowania. W kolejnych krokach ewolucji podczas „crossing-over” gen może się wydłużać lub skracać z jednakowym prawdopodobieństwem, a także mutuje. Okazuje

się, że średnio jednak powoli wydłuża się. Na wejście (przez modyfikację stężenia określonych białek) podawane są „ze środowiska” oscylacje ale z zaburzeniami, na wyjściu żądany jest podobny przebieg – czyli fitness określone jest jako podobieństwo przebiegu stężenia określonych białek do zadanego przebiegu ale bez zakłóceń. Jak się okazało w tym eksperymencie symulacyjnym, stosującym populację 250 osobników przez 250 pokoleń, wystarczyło 5 genów do uzyskania dobrych przebiegów periodycznych na wyjściu, działających poprawnie w obszarach zaburzonego wejścia. Oczywiście, dla większej pewności wyników każdy konkretny eksperyment o określonych typach przebiegów na wejściu powtarzany był 10 razy.

PODSUMOWANIE

Algorytmy ewolucyjne są abstrakcyjną esencją mechanizmów darwinowskich, pozwalającą na precyzyjne zdanie sobie sprawy ze znaczenia każdego ich elementu. Oprócz samego mechanizmu Darwin wskazał ich interpretację biologiczną. Obecnie algorytmy ewolucyjne i zbliżona dziedzina – ewolucja sieci złożonych, badane i używane są nie przez biologów, przez co interpretacja biologiczna ich elementów, wielkości i wyników pozostaje w tyle. Widać to na przykładzie podstawowego pojęcia w tym obszarze jakim jest fitness. Biologia może dużo zyskać z głębszej analizy interpretacji tego pojęcia w kontekście algorytmów ewolucyjnych.

Szczególnie w przypadku ewolucji sieci złożonych, gdzie interpretacja biologiczna jest znacznie trudniejsza i bogatsza, dostrzeżone już kontrowersje interpretacyjne wymagają dalszych badań. Tu np. modele regulacji genowych tego „biologicznego” typu wraz z odtwarzanymi rzeczywistymi sieciami pozwolą na miarodajne sprawdzenie roli ujemnych sprzężeń zwrotnych i jednoznaczne stwierdzenie stopnia poprawności wniosków z modelu Kauffmana podkopujących wyłączność mechanizmów darwinowskich w obszarach źródeł stabilności i obserwowanego uporządkowania organizmów żywych.

EVOLUTIONARY AND GENETIC ALGORITHMS, EVOLUTION OF COMPLEX NETWORKS AND GENE REGULATORY MODELS VERSUS DARWINIAN MECHANISMS

Summary

Darwinian mechanism can be clearly understood and its power can be easily observed in exact form of ‘Evolutionary Algorithms’. In the first step a simplest algorithm is considered (without ‘crossing-over’) to make clear definitions. Fitness is defined as the similarity to an arbitrarily chosen ideal. This method substitutes simulation of long period of object life for measuring its death probability. Typically much more powerful is Genetic Algorithm. It contains additionally a crossing-over mechanism, which allows a set of evolving objects to be a magazine of alternative, simultaneously collected properties. As optimization algorithm, it is typically faster than algorithms without crossing-over.

Basic effects of evolution of complex networks, which can describe living objects, are dis-

cussed using Kauffman (Boolean) networks. The best parameters of network allowing adaptive evolution can be found near transition between order and chaos; Kauffman formulated the hypothesis known as ‘life on the edge of chaos’. Kauffman suggests that spontaneous order (the biggest in this area) is a large part of observed order in living objects. This Kauffman’s hypothesis is currently the strongest attack on area explored by Darwinian mechanisms. Although Kauffman’s hypothesis for his gene regulatory model is experimentally confirmed by measuring stability, I am convinced that negative feedbacks are not taken into account sufficiently. For comparison, the gene regulatory model based on Banzhaf’s idea is shortly described.

LITERATURA

- ALBERT R., BARABÁSI A.-L., 2002. *Statistical mechanics of complex networks*. Rev. Modern Phys. 74, 47–97.
- BARABÁSI A.-L., BONABEAU E., 2003. *Scale-Free Networks*. Sci. Am. 288, 60–9.
- BANZHAF W., 2003. *On the Dynamics of an Artificial Regulatory Network*. [W:] *Advances in Artificial Life, 7th European Conference, ECAL'03*. Lecture Notes in Artificial Intelligence 2801, Springer, 217–227.
- GECOW A., 2008a. *Structural Tendencies – effects of adaptive evolution of complex (chaotic) systems*. Int. J. Mod. Phys. C 19, 647–664.
- GECOW A., 2008b. *The purposeful information. On the difference between natural and artificial life*. Dialogue & Universalism 18, 191–206.
- GECOW A., 2009. *A simplified algorithm for statistical investigation of damage spreading*. BICS 5–7 Nov. 08 Tg. Mures Romania AIP Conf. Proc 1117, 133–141.
- GOLDBERG D. E., 1995. *Algorytmy genetyczne i ich zastosowanie*. Wydawnictwa Naukowo-Techniczne, Warszawa.
- HUGHES T. R., HUGHES T. R., MARTON M. J., JONES A. R., ROBERTS C. J. i współaut., 2000. *Functional discovery via a compendium of expression profiles*. Cell 102, 109–126.
- IGUCHI K., KINOSHITA S. I., YAMADA H., 2007. *Boolean dynamics of Kauffman models with a scale-free network*. J. Theor. Biol. 247, 138–151.
- KAUFFMAN S. A., 1969. *Metabolic stability and epigenesis in randomly constructed genetic nets*. J. Theor. Biol. 22, 437–467.
- KAUFFMAN S. A., 1971. *Gene regulation networks: a theory for their global structure and behaviour*. Curr. Topics Dev. Biol. 6, 145.
- KAUFFMAN S. A., 1993. *The Origins of Order: Self-Organization and Selection in Evolution*. Oxford University Press, New York.
- KNABE J. F., NEHANIV C. L., SCHILSTRA M. J., QUICK T., 2006. *Evolving Biological Clocks using Genetic Regulatory Networks*. [W:] *Proceedings of the Artificial Life X Conference (Alife 10)*. ROCHA L. M., YAEGER L. S., BEDAU M. A., FLOREANO D., GOLDSTONE R. L., VESPIGNANI A. (red.). MIT Press, 15–21.
- KORONA R., 1998. *Mechanizmy ewolucyjne utrzymujące rozmnażanie płciowe*. Kosmos 47, 163–173.
- KORZENIEWSKI B., 2001. *Cybernetic formulation of the definition of life*. J. Theor. Biol. 209, 275–286.
- KORZENIEWSKI B., 2005. *Confrontation of the cybernetic definition of living individual with the real world*. Acta Biotheor. 53, 1–28.
- LUQUE B., BALLESTEROS F. J., 2004. *Random walk networks*. Physica A 342, 207–213.
- RÁMÓ P., KESSELI J., YLI-HARJA O., 2006. *Perturbation avalanches and criticality in gene regulatory networks*. J. Theor. Biol. 242, 164–170.
- SERRA R., VILLANI M., SEMERIA A., 2004. *Genetic network models and statistical properties of gene expression data in knock-out experiments*. J. Theor. Biol. 227, 149–157.
- SERRA R., VILLANI M., GRAUDENZI A., KAUFFMAN S. A., 2007. *Why a simple model of genetic regulatory networks describes the distribution of avalanches in gene expression data*. J. Theor. Biol. 246, 449–460.
- TEMPCZYK M., 1998. *Teoria chaosu a filozofia*. Wydawnictwo CiS, Warszawa
- THOM R., 1970. *Topological models in biology*. [W:] *Towards a Theoretical Biology*. Drafis. WADDINGTON C. H. (red.). Aldine, Chicago.
- WAGNER A., 2001. *Estimating coarse gene network structure from large-scale gene perturbation data*. Santa Fe Institute Working Paper, 01-09-051.
- WŁODARCZYK M., 1998. *Wymiana materiału genetycznego i rekombinacja u bakterii*. Kosmos 47, 137–146.