



Uogólniony Crowding dla algorytmów genetycznych

Severino F. Galan
Wydział Sztucznej Inteligencji
National Distance University of
Spain C/ Juan del Rosal 16
28040 Madryt, Hiszpania
seve@dia.uned.es

Ole J. Mengshoel
Carnegie Mellon University
NASA-Ames Research
Center Mail Stop 263-3, P.O.
Box 1 Moffett Field, CA 94035
ole.mengshoel@sv.cmu.edu

ABSTRAKT

Tłoczenie jest techniką stosowaną w algorytmach genetycznych w celu wstępnego zapewnienia różnorodności w populacji i zapobiegania przedwczesnej konwergencji do lokalnych wartości optymalnych. Polega ona na parowaniu każdego osobnika poza populacją z podobnym osobnikiem w bieżącej populacji (*faza parowania*) i decydowaniu, który z nich pozostanie w populacji (*faza zastępowania*). Niniejsza praca koncentruje się na fazie zastępowania, która zwykle jest przeprowadzana za pomocą jednego z trzech następujących podejść: Deterministycznego, Probabilistycznego i Symulowanego Analizy. Podejścia te wiążą się z pewnymi ograniczeniami dotyczącymi sposobu przeprowadzania zastępowania. Z jednej strony, dwie pierwsze stosują tę samą presję selekcyjną niezależnie od rozwiązywanego problemu lub etapu algorytmu genetycznego. Z drugiej strony, trzeci z nich nie stosuje jednolitej presji selektywnej w stosunku do wszystkich osobników w populacji, co sprawia, że kontrola presji selektywnej na przestrzeni pokoleń jest nieco utrudniona. Niniejsza praca przedstawia podejście *Generalized Crowding*, które pozwala w prosty sposób kontrolować presję selektywną w fazie zastępowania, przewyższając w ten sposób ograniczenia innych podejść. Co więcej, zrozumienie istniejących podejść zostało znacznie poprawione, ponieważ zarówno Deterministyczne, jak i Probabilistyczne Tłoczenie okazują się być szczególnymi przypadkami Uogólnionego Tłoczenia. Ponadto, parametr temperatury używany w symulowanym wyżarzaniu został zastąpiony parametrem zwanym *współczynnikiem skalowania*, który kontroluje zastosowaną selektywną presję. Analiza teoretyczna z wykorzystaniem łańcuchów Markowa i ocena empiryczna z wykorzystaniem sieci bayesowskich pokazują potencjał tego nowatorskiego podejścia Generalized Crowding.

niniejszą informacją i pełnym cytatem na pierwszej stronie. Kopiowanie w inny sposób, ponowne publikowanie, umieszczanie na serwerach lub redystrybucja na listach wymaga uprzedniej specjalnej zgody i/lub opłaty.

GECCO'10, 7-11 lipca 2010, Portland,
Oregon, USA. Copyright 2010 ACM 978-
1-4503-0072-8/10/07 ...\$10.00.

Kategorie i deskryptory tematów

G.3 [Prawdopodobieństwo i statystyka]: Procesy Markowa;
Algorytmy probabilistyczne (w tym Monte Carlo); I.2.8
[Sztuczna inteligencja]: Rozwiązywanie problemów, metody
sterowania i wyszukiwanie - *metody heurystyczne*

Zezwala się na tworzenie cyfrowych lub papierowych kopii całości lub części tej pracy do użytku osobistego lub szkolnego bez opłat, pod warunkiem, że kopie nie będą wykonywane ani rozpowszechniane w celu osiągnięcia zysku lub korzyści komercyjnej oraz że kopie będą opatrzone

Warunki ogólne

Algorytmy

Słowa kluczowe

Algorytmy genetyczne, niching, deterministyczny tŁok, probabilistyczny tŁok, analiza łańcucha Markowa, sieci bayesowskie, eksperymenty

1. WPROWADZENIE

Algorytmy genetyczne (GA) [6, 4] wykorzystują stochastyczne metody wyszukiwania oparte na naturalnej ewolucji w celu rozwiązywania problemów optymalizacyjnych w dziedzinach takich jak projektowanie, uczenie się lub planowanie. Podstawą działania GA są dwa procesy: wariacja (rekombinacja i mutacja) oraz selekcja. Podczas gdy pierwszy z nich sprzyja różnorodności i nowości, drugi faworyzuje jakość. Celem GA jest znalezienie optymalnego (lub zbliżonego do optymalnego) rozwiązania lub zidentyfikowanie zestawu rozwiązań o wysokim dopasowaniu. Przedwczesna zbieżność do lokalnych optymalnych rozwiązań jest jedną z najczęstszych trudności pojawiających się przy stosowaniu GA do złożonych problemów. Wiąże się ona z utratą różnorodności w populacji; jednakże zbyt duża różnorodność populacji może prowadzić do dramatycznego pogorszenia wydajności GA. Dlatego też ważną kwestią w badaniach i praktyce GA jest kompromis pomiędzy eksploatacją najlepszych osobników a eksploracją alternatywnych obszarów przestrzeni poszukiwań.

TŁoczenie zostało wprowadzone przez De Jonga [2] jako technika zachowania różnorodności populacji i zapobiegania przedwczesnej konwergencji. TŁoczenie jest stosowane na etapie wyboru przetrwania w GA w celu podjęcia decyzji, które osobniki spośród tych w bieżącej populacji i ich potomstwo przejdą do następnego pokolenia. Crowding składa się z dwóch głównych faz: parowania i zastępowania. W fazie parowania osobniki potomne są parowane z osobnikami w bieżącej populacji zgodnie z metryką podobieństwa. W fazie zastępowania dla każdej pary osobników podejmowana jest decyzja, który z nich pozostanie w populacji. Przegląd podejść crowdingowych dla GA można znaleźć w [13].

W zależności od tego, w jaki sposób przeprowadzana jest faza zastępowania, istnieją trzy główne typy zatłoczenia: Deterministyczny [8, 9], Probabilistyczny [12, 11] i oparty na symulowanym wyżarzaniu [10]. Deterministyczny Crowding wybiera najlepiej dopasowanego osobnika w każdej parze w fazie zastępowania. TŁoczenie probabilistyczne wybiera ocalałego osobnika dla każdej pary w oparciu o formułę probabilistyczną, która bierze pod uwagę kondycję. Wreszcie, crowding inspirowany symulowanym wyżarzaniem wykorzystuje dobrze znane reguły, takie jak Metropolis lub Boltzmann, które zawierają parametr temperatury w fazie zastępowania.

Projekt reguły zastępowania ma ogromny wpływ na wydajność tłumienia. Deterministic Crowding rozwija strategię zastępowania opartą na wykorzystaniu. Może to być wadą, ponieważ eksploatacyjne metody w GA często prowadzą do przedwczesnej zbieżności. Probabilistic Crowding promuje eksplorację alternatywnych (mniej dopasowanych) rozwiązań. Stopień eksploracji jest definiowany za pomocą formuły probabilistycznej, która nie jest zmieniana w trakcie wykonywania GA. Jednak kompromis między eksploracją a eksploatacją można osiągnąć tylko wtedy, gdy selektywna presja może być dostosowana do rozwiązywanego problemu i etapu GA. Tłoczenie oparte na regułach zastępowania symulowanego wyżarzania pozwala na zmianę stopnia eksploracji za pomocą parametru temperatury. Pomimo tego, biorąc pod uwagę temperaturę i populację, jednolita presja selektywna nie jest stosowana do wszystkich osobników w populacji, jak pokazano w sekcji 2.2. Sprawia to, że kontrola presji selektywnej na pokolenia jest trudniejsza niż być może powinna. Jedną z motywacji do badań opisanych w niniejszym artykule było znalezienie rozwiązania dla trzech ograniczeń wymienionych powyżej.

Niniejsza praca koncentruje się na fazie zastępowania tłum i przedstawia podejście *Generalized Crowding*, które pozwala w prosty sposób kontrolować selektywną presję. Uogólniony Crowding jest inspirowany Probabilistycznym Crowdingiem, ale różni się od niego tym, że stopień eksploracji można zmienić za pomocą parametru zwanego *współczynnikiem skalowania*. Zarówno Deterministic, jak i Probabilistic Crowding okazują się być specjalnymi przypadkami Generalized Crowding, uzyskanymi poprzez przypisanie współczynnikowi skalowania wartości odpowiednio 0 i 1. Określona wartość współczynnika skalowania zapewnia populacji sposób na zastosowanie jednolitej selektywnej presji na wszystkie jej osobniki. Podobnie jak w przypadku temperatury, współczynnik skalowania może być obniżany na przestrzeni pokoleń zgodnie z ustalonym harmonogramem.

Pozostała część tego artykułu jest zorganizowana w następujący sposób. Najpierw przedstawiamy wstępne informacje, w tym wcześniejsze badania nad tłumem w GA. W następnej sekcji przedstawiamy nasze podejście Generalized Crowding, które opiera się na wykorzystaniu parametrycznego współczynnika skalowania. Następnie analizujemy nasze nowatorskie podejście za pomocą (dyskretnego czasu, dyskretnego przestrzeni stanów) Łańcuchy Markowa. Następnie przedstawiamy wyniki eksperymentalne dla

wykorzystanie Generalized Crowding do oszacowania najbardziej prawdopodobnego wyjaśnienia w sieciach bayesowskich (BN), podczas gdy ostatnia sekcja podsumowuje i nakreśla obszary przyszłych badań.

2. PODEJŚCIE CROWDINGOWE

Najpierw wprowadzimy kilka definicji: p_c - prawdopodobieństwo krzyżowania; p_m - prawdopodobieństwo mutacji; M - wielkość populacji; a S jest rozmiarem rodziny [13]. Bez utraty ogólności przyjmujemy maksymalizację funkcji fitness $f: \{0, 1\}^n \rightarrow \mathbb{R}$ w tej sekcji.

Oryginalny schemat tłumienia opracowany przez De Jonga w latach siedemdziesiątych [2] polega na losowym wyborze dla każdego potomstwa γ osobników z bieżącej populacji. Potomstwo zastąpi najbardziej podobnego wybranego osobnika. Parametr γ jest znany jako *współczynnik zatłoczenia* i zwykle stosuje się $\gamma = 2$.

W warunkach zatłoczenia każdy osobnik w populacji staje się rodzicem).

2. Z prawdopodobieństwem p_c , rodzice w każdej parze (p_1, p_2) są rekombinowani. Dwoje powstałych w ten sposób potomków (c_1, c_2) jest mutowanych z prawdopodobieństwem p_m .
3. Każde dziecko konkuruje z jednym ze swoich rodziców o włączenie do populacji następnego pokolenia. Niech $d(i_1, i_2)$ oznacza odległość między dwoma osobnikami, i_1 i i_2 :

Jeśli $d(p_1, c_1) + d(p_2, c_2) < d(p_1, c_2) + d(p_2, c_1)$

$p_1 \leftarrow$ zwycięzca rywalizacji pomiędzy p_1 i c_1 p_2

\leftarrow zwycięzca rywalizacji pomiędzy p_2 i c_2

inny

$p_1 \leftarrow$ zwycięzca rywalizacji pomiędzy p_1 i c_2 p_2

\leftarrow zwycięzca rywalizacji pomiędzy p_2 i c_1

W ramach tego schematu każde potomstwo ma tendencję do konkurowania o przetrwanie ze swoim najbardziej podobnym rodzicem. Istnieją inne warianty, które wybierają więcej niż dwoje rodziców i dzieci przed zastosowaniem metryki podobieństwa [13, sekcja 4.2]. Pomysł ten był podstawą wielu szeroko stosowanych nowoczesnych podejść do tłumienia [8, 9, 12, 11, 10]. Jedną z kluczowych różnic między tymi podejściami jest reguła stosowana do wyłonienia zwycięzcy każdego konkursu, co omówimy poniżej.

2.1 Konkurencja i wymiana

W zatłoczeniu sposób, w jaki odbywa się konkurencja między rodzicem p a dzieckiem c , jest definiowany za pomocą tak zwanej *reguły zastępowania*. Omówimy teraz cztery z najczęściej stosowanych reguł zastępowania: *zastępowanie deterministyczne*, *zastępowanie probabilistyczne*, *zastępowanie Boltzmanna* i *zastępowanie Metropolisa*. Dwie ostatnie opierają się na zastosowaniu idei symulowanego wyżarzania [7] do tłumienia.

W metodzie *Deterministic Crowding* [8, 9] zwycięzca rywalizacji między rodzicem p a dzieckiem c jest to, które ma wyższą kondycję. Niech p_c oznacza prawdopodobieństwo, że dziecko c zastąpi rodzica p w populacji. Prawdopodobieństwo to można przedstawić w następujący sposób dla deterministycznego zastępowania:

$$p_c = \begin{cases} 1 & \text{jeśli } f(c) > f(p) \\ 0,5 & \text{jeśli } f(c) = f(p) \\ 0 & \text{jeśli } f(c) < f(p) \end{cases}$$

W przeciwieństwie do Deterministic Crowding, *Probabilistic Crowding* [12, 11] wykorzystuje niedeterministyczną regułę do ustalenia zwycięzcy rywalizacji między rodzicem p a dzieckiem c . Prawdopodobieństwo, że c zastąpi p w populacji jest następujące:

$$p_c = \frac{f(c)}{f(c) + f(p)}. \quad (1)$$

Boltzmann Crowding [10] opiera się na dobrze znanej metodzie Sim-

Metoda ulated Annealing, zaimplementowana z regułą akceptacji Boltzmanna [1, 5]. Zgodnie z tą metodą zastępowania, p_c przyjmuje wyrażenie:

Schemat De Jonga został nieco zmieniony w latach dziewięćdziesiątych przez Mahfouda [8, 9]. Ponieważ potomstwo może być podobne do jego rodziców, można zastosować następujący schemat, aby skutecznie zachować różnorodności w populacji:

1. Osobniki w bieżącej populacji są losowo łączone w pary. (Ponieważ selekcja rodziców nie jest zwykle stosowana

$$P_c = \frac{1}{1 + \exp\left(\frac{f(p) - f(c)}{T}\right)} \quad (2)$$

gdzie T jest temperaturą w bieżącej generacji. Temperatura jest zwykle obniżana w kolejnych generacjach za pomocą harmonogramu chłodzenia. Dwa typowe harmonogramy chłodzenia polegają na powtarzaniu przez generację jednej z następujących operacji: mnożenie bieżącej temperatury przez stałą

nizszą od jedności lub odejmując stałą wartość od bieżącej temperatury. Niektórzy autorzy stosowali również stałą temperaturę.

Like Boltzmann Crowding, *Metropolis Crowding* [10] applies Simulated Annealing in the replacement phase. However, the Metropolis acceptance rule [14] is used instead of the Boltzmann rule. The probability that child c replaces parent p in the population is calculated as follows:

$$P_c = \begin{cases} 1 & \text{jeśli } f(c) \geq f(p) \\ \exp\left(-\frac{f(p)-f(c)}{T}\right) & \text{if } f(c) < f(p) \end{cases}, \quad (3)$$

gdzie T jest temperaturą w bieżącej generacji. The można zastosować taki sam harmonogram chłodzenia jak w przypadku Boltzmann Crowding; alternatywnie można zastosować stałą temperaturę.

2.2 Dyskusja

Prawdopodobieństwo P_c , że dziecko c zastąpi rodzica p w populacji, jest obliczane na różne sposoby w zależności od zastosowanej konkretnej reguły zastępowania. Podczas gdy w regułach opartych na symulowanym wyżarzaniu, zastępowaniu Boltzmann i zastępowaniu Metropolis, P_c jest funkcją $f(p) - f(c)$, jak pokazano w równaniach 2 i 3, w zastępowaniu probabilistycznym P_c jest funkcją $f(p)/f(c)$, co można zobaczyć, dzieląc licznik i mianownik w równaniu 1 przez $f(c)$.

Biorąc pod uwagę rodzica p z kondycją $f(p)$ i dziecko c z kondycją $f(c)$, całkiem różne presje selektywne uzyskuje się ogólnie z $P_c = g_1(f(p) - f(c))$ i $P_c = g_2(f(p)/f(c))$, gdzie g_1 i g_2 są funkcjami rzeczywistymi. Na przykład, niech s_1 i s_2 będą następującymi zbiorami par rodzic-dziecko: $s_1 \equiv \{(f(p_1) = 2, f(c_1) = 1), (f(p_2) = 12, f(c_2) = 11), (f(p_3) = 102, f(c_3) = 101)\}$ i $s_2 \equiv \{(f(p_1) = 2, f(c_1) = 1), (f(p_2) = 20, f(c_2) = 10), (f(p_3) = 200, f(c_3) = 100)\}$.

Podczas gdy zastosowanie $P_c = g_1(f(p) - f(c))$ dałoby stałe wartości P_c dla elementów w s_1 ($f(p) - f(c) = 1$ dla każdej pary w s_1), zastosowanie $P_c = g_2(f(p)/f(c))$ dałoby stałą wartość P_c dla elementów w s_1 ($f(p) - f(c) = 1$ dla każdej pary w s_1).

powoduje niestale wartości P_c dla s_1 . W przeciwnym razie mają miejsce dla s_2 . Praca ta koncentruje się na wykorzystaniu funkcji

$f(p)/f(c)$ w fazie wymiany, ze względu na fakt, że presja selektywna jest częściej stosowana w GA poprzez niwelując tę opcję zamiast opcji wykorzystującej funkcję $f(p) - f(c)$. Na przykład, *proporcjonalny wybór fitness* [6] jest szeroko stosowaną metodą selekcji w GA, która jest matematycznie równoważna Probabilistic Crowding: Biorąc pod uwagę dwa osobniki p i c , ten drugi jest wybierany z prawdopodobieństwem

$\frac{f(c)}{f(c)+f(p)}$ w ramach selekcji proporcjonalnej, która jest równoważna formule stosowanej w zastępowaniu probabilistycznym.

Interesujące jest przeanalizowanie wartości, które symulowane Annealing wygenerowałby dla P_c w przypadku par w s_2 :

$$P_{c1} = \exp\left(-\frac{2-1}{1}\right) = 0.36788$$

$$P_{c2} = \exp\left(-\frac{20-10}{1}\right) = 4.54 \times 10^{-5}$$

$$P_{c3} = \exp\left(-\frac{200-100}{1}\right) = 3.7201 \times 10^{-44}$$

gdzie $T = 1$ i zastosowano zastępowanie Metropolis. Jak pokazano w tabeli 1, podczas gdy w przypadku

Zestaw	$f(p_i)$	$f(c_i)$	$\frac{f(p_i)}{f(c_i)}$	$\frac{f(p_i)}{f(c_i)}$	$P_{c,PR}$	$P_{c,MR}$
s1	2	1	1	2	0.333	0.368
	12	11		1.09	0.478	
	102	101		1.01	0.498	
s2	2	1	1	2	0.333	0.368
	20	10				4.54×10^{-5}
	200	100				3.72×10^{-44}

Tabela 1: Wartości P_c dla par w s_1 i s_2 w warunkach zastępowanie probabilistyczne ($P_{c,PR}$) i ponowne umieszczanie Metropolis ($P_{c,MR}$). $P_{c,PR}$ jest niezmiennie dla s_2 , podczas gdy $P_{c,MR}$ jest niezmiennie dla s_1 .

Aby stworzyć regułę zastępowania z zaletami zarówno zastępowania probabilistycznego (gdzie P_c jest funkcją $f(p)/f(c)$ dla wszystkich osobników w bieżącej populacji), jak i zastępowania symulowanego wyżarzania (gdzie P_c można kontrolować przez pokolenia za pomocą parametru, w tym przypadku temperatury), następna sekcja wprowadza nową metodę tłumienia, którą nazywamy *Generalized Crowding*.

3. UOGÓLNIONY TŁOK

Opierając się na schematach omówionych w poprzedniej sekcji, Generalized Crowding składa się z fazy parowania i fazy ponownego umieszczania. Nowością jest to, że faza zastępowania opiera się na wykorzystaniu współczynnika skalowania ϕ , co pozwala nam na przeprowadzenie szerokiego zakresu reguł zastępowania tylko poprzez dostosowanie ϕ , w porównaniu do konieczności radzenia sobie z potencjalnie bardzo różnymi algorytmami zastępowania.

W Generalized Crowding zwycięzca rywalizacji między rodzicem p i dzieckiem c jest ustalany za pomocą:

$$P_c = \begin{cases} \frac{f(c)}{f(c)+\phi \times f(p)} & \text{jeśli } f(c) > f(p) \\ 0.5 & \text{jeśli } f(c) = f(p) \\ \frac{f(p)}{\phi \times f(c)+f(p)} & \text{if } f(c) < f(p) \end{cases}, \quad (4)$$

zastępowania probabilistycznego P_{ci} byłoby równe $1/3$ dla $i \in \{1, 2, 3\}$, w przypadku zastępowania Metropolis uzyskuje się niejednorodny zestaw wartości P_c . Na przykład $P_{c1} = 0.36788$ jest podobne do zastosowania podstawienia probabilistycznego, podczas gdy $P_{c3} = 3.7201 \times 10^{-44}$ jest prawie takie samo jak zastosowanie podstawienia deterministycznego.

gdzie P_c to prawdopodobieństwo, że dziecko c zastąpi rodzica p w populacji, f to funkcja fitness, która ma zostać zmaksymalizowana, a $\varphi \in \mathbb{R}^+ \cup \{0\}$ oznacza parametr o nazwie *współczynnik skalowania*. Kluczową ideą Generalized Crowding jest to, że skalowanie kondycji najmniej dopasowanego osobnika spośród p i c odbywa się przed zastosowaniem Probabilistic Crowding. Jak pokazano w (4), jeśli $f(p) < f(c)$, to $f(p)$ jest przekształcane w $\varphi \times f(p)$; w przeciwnym razie $f(c)$ jest przekształcane w $\varphi \times f(p)$. W szczególności, gdy $\varphi = 0$ w równaniu 4, uogólniony Crowding staje się odpowiednikiem Deterministic Crowding. Gdy $0 < \varphi < 1$, możliwe jest, że najmniej dopasowany spośród p i c wygrywa w fazie wymiany. Gdy $\varphi = 1$, Generalized Crowding zamienia się w Probabilistic Crowding. Wreszcie, gdy $\varphi > 1$, prawdopodobieństwo, że najmniej dopasowane spośród p i c wygra jest większe niż w Probabilistic Crowding, co może być korzystne dla zwodniczych lub wysoce multimodalnych problemów.

Aby uniknąć konieczności radzenia sobie z trzema oddzielnymi przypadkami w (4), można użyć funkcji logistycznej $l(x) = 1/(1+e^{-x})$, aby uzyskać:

$$P_c = \frac{\varphi E(f(p)-f(c)) \times f(c)}{\varphi E(f(p)-f(c)) \times f(c) + \varphi E(f(c)-f(p)) \times f(p)} \quad (5)$$

To skalowanie (4) lub (5) różni się od skalowania w symulowanym wyżarzaniu z dwóch głównych powodów:

1. W Generalized Crowding zmienia się albo $f(p)$ albo $f(c)$, podczas gdy w Simulated Annealing zmienia się wielkość $f(p) - f(c)$. Innymi słowy, zgodnie z dyskusją w sekcji 2.2, Generalized Crowding stosuje selektywną presję jako funkcję $f(p)/f(c)$.
2. Operacja skalowania jest prosta w Generalized Crowding, to znaczy jest to tylko mnożenie bez dzielenia lub wykładników. Pozwala to uogólnionemu Crowdingowi na włączenie Deterministycznego Crowdingu i Probabilistycznego Crowdingu jako specjalnych przypadków.

Współczynnik skalowania ϕ pozwala na zastosowanie szerokiego zakresu selektywnych założeń: Im większy jest ϕ , tym bardziej prawdopodobne jest, że lokalnie nieoptymalne części przestrzeni poszukiwań zostaną zbadane. Podobnie do symulowanego wyżarzania, ϕ może być poddane eksploracji i zmianom, który faworyzuje

generacji GA i stopniowo zwiększa stopień wykorzystania najlepszych rozwiązań.

4. ANALIZA ŁAŃCUCHA MARKOWA

W tej sekcji używamy czasu dyskretnego, dyskretniej przestrzeni stanów

Łańcuchy Markowa do analizy naszego podejścia

Generalized Crowding. Najpierw przedstawiamy ogólny przypadek dla N nisz, a następnie rozwiązujemy przypadek dla dwóch nisz przy użyciu ogólnych równań dla N nisz.

4.1 Wiele nisz

Przeprowadzamy teraz analizę, w której każda nisza, w której może znajdować się rodzic, ma odpowiadający jej stan łańcucha Markowa. Istnieją dodatkowe stany łańcucha Markowa, po jednym dla każdej kombinacji wszystkich możliwych lokalizacji niszowych dla par dziecka i rodzica. Formalnie wprowadzamy następujące definicje: p jest rodzicem w bieżącym pokoleniu; p^* jest rodzicem w następnym pokoleniu; c jest dzieckiem w bieżącym pokoleniu; N jest liczbą nisz; oraz¹

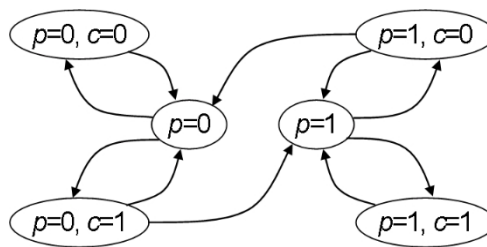
$$\begin{aligned} \Pr([p = i, c = j] \mid [p = i]) &= A_{i,j}, i, j \in \{0, \dots, N - 1\} \\ \Pr([p^* = j] \mid [p = i, c = j]) &= B_{i,j}, i, j \in \{0, \dots, N - 1\}, \end{aligned}$$

gdzie $A_{i,j}$ oznacza prawdopodobieństwo, że dany rodzic p w niszy i jest sparowany z dzieckiem c w niszy j , a $B_{i,j}$ oznacza prawdopodobieństwo, że dany rodzic p w niszy i sparowany z dzieckiem c w niszy j , wynikowy rodzic p^* dla następnego pokolenia znajduje się w niszy j po zastąpieniu.

Przed rozwiązaniem równań równowagi dla $i, j \in \{0, \dots, N - 1\}$ musimy obliczyć następujące parametry:

$$\begin{aligned} \Pr([p^* = j] \mid [p = i]) &= \sum_{k=0}^{N-1} \Pr([p^* = j] \mid [p = i, c = k]) \times \\ &\quad \Pr([p = i, c = k] \mid [p = i]). \end{aligned}$$

Mamy tutaj dwa przypadki: (i) Jeśli $i \neq j$, to $\Pr([p^* = j] \mid [p = i]) =$



Rysunek 1: Rozszerzony łańcuch Markowa (AMC).

Otrzymujemy teraz następujące równania bilansowe:

$$\begin{aligned} N-1 \\ k=\sum 0, k/ \\ N-1 \\ =i \\ \Pr([p^* = k] \mid [p = i]) \Pr([p = i]) = \\ \Pr([p^* = i] \mid [p = k]) \Pr([p = k]), \\ =i \\ i]) = A_{i,j} B_{i,j}; \text{ (ii) jeśli } i = j \text{ to } \Pr([p^* = j] \mid [p = i]) \\ = 1 - \sum_{k \neq i} A_{i,k} B_{i,k}. \end{aligned}$$

¹Zauważ, że podczas gdy $\Pr(A, B \mid B) = \Pr(A \mid B)$ dla zdarzeń A i B , mamy $\Pr([p = i, c = j] \mid [p = i]) \neq \Pr([c = j] \mid [p = i])$. Innymi słowy, $[p = i, c = j]$ jest odrębnym zdarzeniem i nie można go uprościć do $[c = j]$, nawet jeśli warunkujemy $[p = i]$.

ten układ N równań z otrzymanych wzorów

$$\sum_{k=0}^{N-1}$$
w (i) i (ii) dla $\Pr([p^* = j] \mid [p = i])$, możemy otrzymać $\Pr([p = k])$ dla $k \in \{0, \dots, N - 1\}$. Jednak wynikowe wyrażenia są dość złożone dla ogólnego przypadku. Dlatego w tej pracy bardziej szczegółowo przeanalizujemy przypadek dwóch nisz.

4.2 Dwie nisze

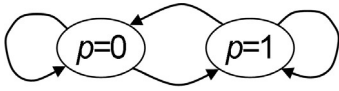
Aby przeanalizować przypadek dwóch nisz, wprowadzamy dwa łańcuchy Markowa. Bardziej szczegółowy rozszerzony łańcuch Markowa, który omawiamy jako pierwszy, jest używany jako pomoc w definiowaniu prawdopodobieństw przejścia w dwustanowym pokoleniowym łańcuchu Markowa, gdzie każdy stan odpowiada niszy.

4.2.1 Rozszerzony łańcuch Markowa

Graf dla rozszerzonego łańcucha Markowa (AMC) pokazano na rysunku 1. Macierz przejścia przedstawiono w tabeli 2; wyraźnie pokazuje ona prawdopodobieństwa przejścia A, A^*, B, B^* i inne. Stany $[p = 0]$ i $[p = 1]$ reprezentują zawartość lokalizacji populacji w określonym czasie; zawiera ona osobnika (rodzica), który znajduje się w niszy 0 (więc $p = 0$) lub niszy 1 (więc $p = 1$). Rozważmy lokalizację, która zawiera osobnika w niszy 1, więc $p = 1$. Na etapie konkurencji (po krzyżowaniu, mutacji i parowaniu) istnieją dwie opcje dla p : albo konkurujące dziecko znajduje się w niszy 0 (więc $c = 0$), albo w niszy 1 (więc $c = 1$). W rezultacie mamy dwa stany, a mianowicie $[p = 1, c = 0]$ i $[p = 1, c = 1]$, i istnieją krawędzie od $[p = 1]$ do każdej z nich. Po rywalizacji pozostaje nam osobnik w niszy 0 lub 1, który staje się rodzicem w następnym pokoleniu, w związku z czym istnieją dwie krawędzie z $[p = 1, c = 0]$ - odpowiednio do $[p = 0]$ i $[p = 1]$ - oraz dwie podobne krawędzie z $[p = 1, c = 1]$. Podobna logika ma zastosowanie do stanów $[p = 0, c = 0]$ i $[p = 0, c = 1]$.

Najpierw rozważymy, w jaki sposób prawdopodobieństwa przejścia zależą od krzyżowania, mutacji i parowania w algorytmie crowdingu (A i A^*). Wprowadzamy następujące definicje: $V(0)$ to objętość niszy 0; $V(1)$ to objętość niszy 1; $A(0)$ to powierzchnia niszy 0; a $A(1)$ to powierzchnia niszy 1.

i dodatkowo mamy $\sum \Pr([p = k]) = 1$.
Rozwiązując



Rysunek 2: Generacyjny łańcuch Markowa (GMC).

Dla A i A^* mamy:

$$A = \frac{c_A \times P_E \times P_M \times A(1)}{S \times V(1)}$$

$$A^* = \frac{c_A' \times P_E \times P_M \times A(0)}{S \times V(0)},$$

gdzie c_A i c_A' są stałymi. Dla uproszczenia zakładamy, że $c_A = c_A'$ w pozostałej części tego artykułu.

Po drugie, rozważamy prawdopodobieństwa, które zależą od zastępowania (B i B^*). Wprowadzamy następujące definicje: $f(0)$ to kondycja dla niszy 0, a $f(1)$ to kondycja dla niszy 1. Korzystając z (5) mamy teraz:

$$B = \frac{f(0)}{f(0) + \phi \cdot E(f(0) - f(1)) - E(f(1) - f(0)) \times f(1)}$$

$$B^* = \frac{f(1)}{f(1) + \phi \cdot E(f(1) - f(0)) - E(f(0) - f(1)) \times f(0)}$$

4.2.2 Generacyjny łańcuch Markowa

Graf pokoleniowego łańcucha Markowa (GMC) pokazano na rysunku 2. Jego dwa stany odpowiadają odpowiednio dwóm stanom o tych samych etykietach w AMC. Cztery główne stany AMC są abstrahowane do czterech prawdopodobieństw przejścia GMC, jak zobaczymy wkrótce. Przygotowując się do określenia macierzy przejścia GMC, wprowadzamy następujące definicje stanów AMC: $0 \equiv [p=0]$; $1 \equiv [p=1]$; $00 \equiv [p=0, c=0]$; $01 \equiv [p=0, c=1]$; $10 \equiv [p=1, c=0]$; oraz $11 \equiv [p=1, c=1]$.

Cztery prawdopodobieństwa przejścia GMC można teraz łatwo wyprowadzić z prawdopodobieństw przejścia AMC:

$$\Pr(0 | 0) = \Pr(0 | 00) \Pr(00 | 0) + \Pr(0 | 01) \Pr(01 | 0) = 1 - B A^*,$$

i w podobny sposób otrzymujemy $\Pr(0 | 1) = B A$, $\Pr(1 | 0) = B A^*$, oraz $\Pr(1 | 1) = 1 - B A$. Zakładając równowagę, łatwo jest teraz uzyskać $\Pr(0) \Pr(1 | 0) = \Pr(1) \Pr(0 | 1)$ i $\Pr(0) + \Pr(1) = 1$, co daje następujące rozkłady stacjonarne:

$$\frac{\Pr(0)}{\Pr(1)} = \frac{1 + A B^*}{1 + A B}$$

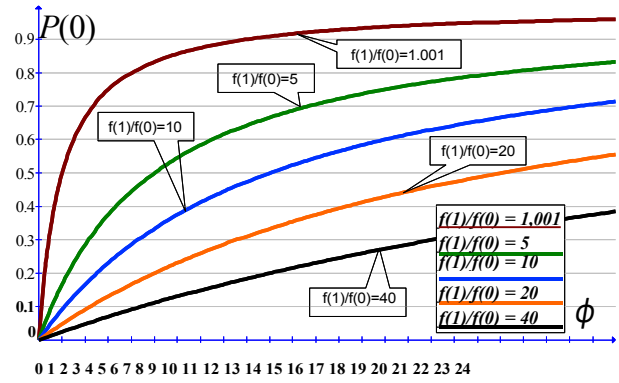
Wprowadzając definicje dla A , A^* , B i B^* , otrzymujemy następujące wyrażenia dla rozkładów stacjonarnych:

$$\Pr(0) = \frac{1}{1 + \frac{A(0)V(1)f(1)}{A(1)V(0)f(0)} \phi \frac{E(f(0)-f(1)) - E(f(1)-f(0))}{E(f(1)-f(0)) - E(f(0)-f(1))}}$$

$$\Pr(1) = \frac{1}{1 + \frac{A(1)V(0)f(0)}{A(0)V(1)f(1)} \phi \frac{E(f(1)-f(0)) - E(f(0)-f(1))}{E(f(1)-f(0)) - E(f(0)-f(1))}}$$

Na podstawie tych dwóch wzorów można wyciągnąć następujące wnioski:

1. Kiedy $\phi = 0$, nisza o najwyższej kondycji otrzymuje całą populację. Jest to Deterministyczne Zatlócenie.



Rysunek 3: Prawdopodobieństwo stacjonarne $\Pr(0)$ (wzdłuż osi y) jako funkcja ϕ (wzdłuż osi x) dla różnych współczynników funkcji fitness $f(1)/f(0)$.

2. Gdy $\phi = 1$, prawdopodobieństwo niszy i wynosi $f(i)/(f(i) + f(j))$, gdzie $i, j \in \{0, 1\}$. To jest Probabilistic Crowding.
3. Gdy $\phi = f(j)/f(i)$, gdzie $f(i) < f(j)$, populacja rozkłada się równo między obie nisze.
4. Gdy $\phi > f(j)/f(i)$, gdzie $f(i) < f(j)$, nisza o mniejszej kondycji otrzymuje więcej populacji niż inna nisza. Wraz ze wzrostem ϕ , jeszcze więcej populacji trafi do niszy o mniejszej kondycji w odniesieniu do drugiej niszy w rozkładzie stacjonarnym.

Przypadek 4 może być korzystny w przypadku problemów multimodalnych, gdzie na przykład przypadek 1 może prowadzić do przedwczesnej zbieżności. Na przykład w eksperymentach z sekcji 5 z siecią bayesowską oznaczoną Alarm0.2, $\phi = 10$ było wartością dającą najlepsze wyniki, gdy $M = 20$.

4.2.3 Populacja

Biorąc pod uwagę, że mamy powyższe wyniki dla jednej lokalizacji populacji, jak można je rozszerzyć na populację o wielkości M ? Używamy M łańcuchów Markowa zamiast jednego łańcucha Markowa. W przypadku jednego łańcucha Markowa i dwóch nisz $\{0, 1\}$, rozkład stacjonarny $\Pr(i)$ dla $i \in \{0, 1\}$ jest zasadniczo próbą Bernoulliego dla niszy i . Tak więc dla M lokalizacji populacji (łańcuchów Markowa) mamy rozkład dwumianowy. W związku z tym średnia wynosi $M \Pr(i)$, a wariancja $M \Pr(i)(1 - \Pr(i))$. Średnia to oczekiwana liczba osobników, które spodziewamy się zobaczyć w niszy. W porównaniu z poprzednimi wynikami dotyczącymi wielkości populacji [13], mamy teraz bardziej ogólne wyrażenia obejmujące ϕ dla rozkładów stacjonarnych.

4.2.4 Przykład

Rysunek 3, na którym rozważamy dwie nisze 0 i 1, pokazuje ϕ wzdłuż osi x i stacjonarne prawdopodobieństwo dla niszy 0, $\Pr(0)$, wzdłuż osi y . Zakładamy, że $f(0) < f(1)$ i narysowaliśmy wykresy dla różnych wartości $f(1)/f(0)$. W tym przypadku $f(0) < f(1)$ mamy:

$$\Pr(0) = \frac{1}{1 + \frac{f(1)}{f(0)}}$$

$$\varphi f^{(0)}$$

które otrzymaliśmy przez (i) umieszczenie $V(0) = V(1)$ i $A(0) = A(1)$ oraz (ii) nie używając funkcji logistycznej, z powodu

	$p = 0$	$p = 1$	$p = 0, c = 0$	$p = 0, c = 1$	$p = 1, c = 0$	$p = 1, c = 1$
$p = 0$	0	0	$1 - A^*$	A^*	0	0
$p = 1$	0	0	0	0	A	$1 - A$
$p = 0, c = 0$	1	0	0	0	0	0
$p = 0, c = 1$	$1 - B^*$	B^*	0	0	0	0
$p = 1, c = 0$	B	$1 - B$	0	0	0	0
$p = 1, c = 1$	0	1	0	0	0	0

Tabela 2: Macierz przejścia dla AMC przedstawiona na rysunku 1.

do faktu, że $\text{Pr}(0)$ nie może być wyrażone jako funkcja $f(1)/f(0)$, jeśli używana jest funkcja logistyczna. Na podstawie tych wykresów można poczynić następujące obserwacje:

- Im większa wartość ϕ , tym większa część populacji przechodzi do mniej dopasowanej niszy 0.
- Dla niskich wartości ϕ ($0 \leq \phi \leq 1$), im bardziej dopasowana nisza 1, tym większy odsetek populacji.
- Jeśli ϕ jest utrzymywane na stałym poziomie, im wyższy jest stosunek $f(1)/f(0)$, tym więcej populacji trafia do bardziej dopasowanej niszy 1.
- Chociaż $f(0) < f(1)$, jeśli $\phi = f(1)/f(0)$, to $\text{Pr}(0) = \text{Pr}(1) = 0,5$, a populacja jest równomiernie rozłożona między nisze 0 i 1. Na przykład wartości ϕ , przy których krzywe osiągają $\text{Pr}(0) = 0,5$, różnią się w następujący sposób: krzywa $f(1)/f(0) = 5$ przecina $\text{Pr}(0) = 0,5$ przy $\phi = 5$, a krzywa $f(1)/f(0) = 40$ przecina $\text{Pr}(0) = 0,5$ przy $\phi = 40$.
- Dla wystarczająco wysokich wartości ϕ , nawet jeśli $f(0) < f(1)$, niszę 0 otrzymuje większość populacji.

Rozważmy pytanie: Dlaczego wartości $\phi > 1$ działają tak dobrze w niektórych przypadkach? Stawiamy hipotezę, że w wysocze wielo-modalnych problemach stopień zwodniczości jest ogólnie wysoki i w konsekwencji, lokalnie, osobnik o wyższym dopasowaniu może być dalej od globalnego optimum niż inny, mniej dopasowany osobnik.

5. EKSPERYMENTY

W tych eksperymentach badamy, w jaki sposób wydajność Generalized Crowding zależy od współczynnika skalowania ϕ . Do celów eksperymentalnych rozważamy szacowanie najbardziej prawdopodobnych wyjaśnień (MPE) w BN, trudnym obliczeniowo problemie interesującym w wielu zastosowaniach, w tym w rozpoznawaniu obrazów, diagnostyce i dekodowaniu korekcji błędów.

5.1 Warianty alarmowej sieci bayesowskiej

W tych eksperymentach wykorzystaliśmy warianty dobrze znanego Alarm BN do zbadania wydajności Generalized Crowding. Alarm BN ma 37 węzłów i 752 wartości CPT (żadna nie jest zerowa).² Losowo wprowadziliśmy zera do CPT BN w celu zróżnicowania charakterystyki przestrzeni wyszukiwania. W szczególności, zera zostały wprowadzone z prawdopodobieństwem $\rho \in \{0.0, 0.1, 0.2, \dots\}$ [3], co odzwierciedla notacja *Alarm*. *Alarm* BN nie jest tym samym co Alarm BN. Mają tę samą strukturę grafu, kardynalność węzłów i liczbę wartości CPT, ale wartości CPT są generowane w sposób losowy dla *Alarm* [3]. W przypadku

Alarm liczba zerowych wartości CPT wynosi około $\rho \times 100\%$. W eksperymentach z tymi

²Szczegółowe informacje na temat alarmu można znaleźć na stronie <http://compbio.cs.huji.ac.il/Repository/Datasets/alarm/alarm.htm>.

Alarm BNs zmienialiśmy φ dla Generalized Crowding i testowaliśmy $\varphi = 0$ (Deterministic Crowding), $\varphi = 0,5$, $\varphi = 1$ (Probabilistic Crowding), $\varphi = 10$ i $\varphi = 100$. W tych dwóch ostatnich przypadkach częściej niż nie gorsze dopasowanie między rodzicem a dzieckiem wygrywa i zaskakująco odkryliśmy, że $\varphi = 10$ przewyższało $\varphi = 0$ i $\varphi = 1$ w niektórych eksperymentach. Ponadto zmieniliśmy rozmiar populacji M .

Zastosowano sparametryzowane ujednolicone krzyżowanie z $p_M = 1$. W przypadku sieci BN z N węzłami użyliśmy $p_M = 1/N$, a mutacja polegała na równomiernej losowej zmianie stanu węzła. Wyniki eksperymentalne są uśrednione dla 350 przebiegów.

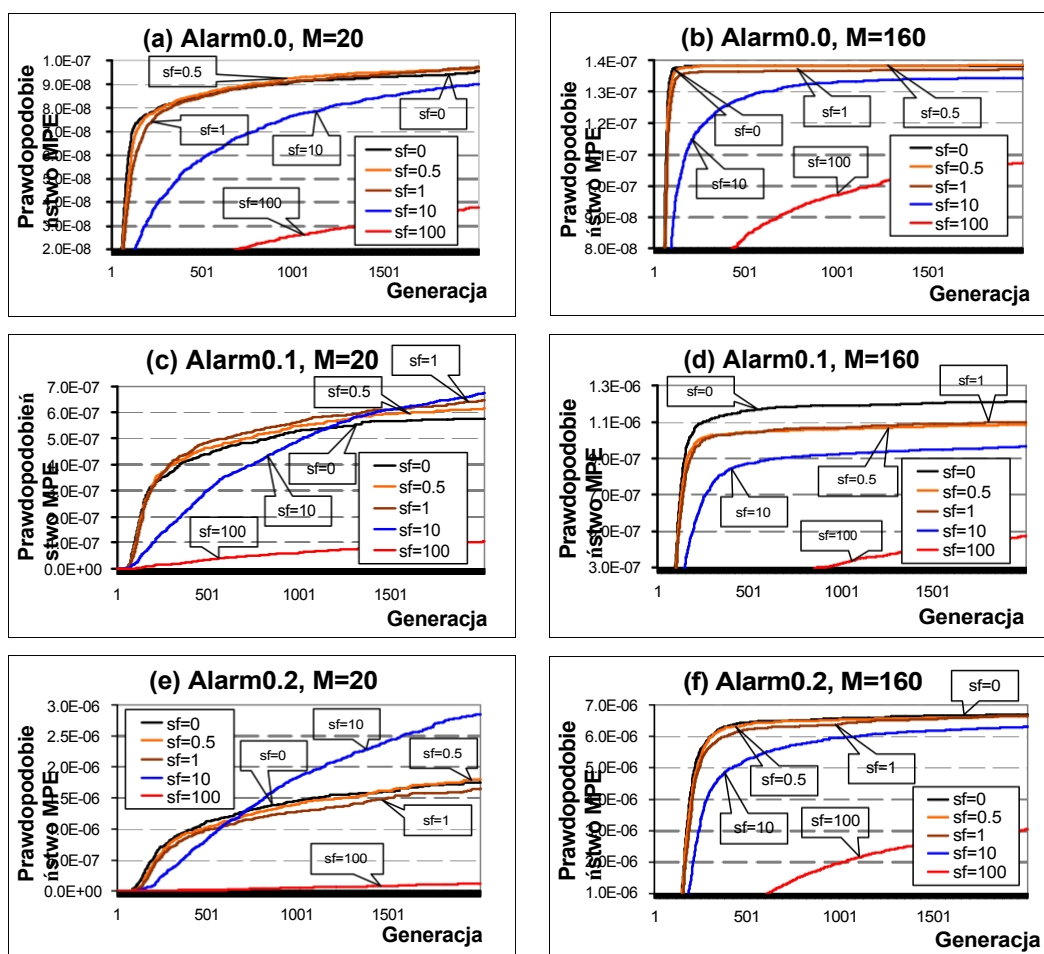
Wyniki eksperymentów podsumowano na rysunku 4. Na podstawie eksperymentów wyciągnęliśmy następujące wnioski. Dla *Alarm*0.0, im wyższy współczynnik skalowania φ , tym gorsza wydajność dla wszystkich rozmiarów populacji³ $M \in \{20, 40, 80, 160\}$. Dla *Alarm*0.1 i przy $M \in \{40, 80, 160\}$ obowiązuje to samo, co dla *Alarm*0.0. Istnieje jednak ważna zmiana dla $M = 20$ (patrz rysunek 4(c)): im wyższa wartość φ , tym lepsza wydajność przy generacji 2000 (z wyjątkiem $\varphi = 100$). To, co powiedzieliśmy dla *Alarm* 0.1, ma ponownie zastosowanie do *Alarm* 0.2, ale z większą intensywnością i zwrótem, że $M = 40$ jest teraz bardziej podobny do *Alarm* 0.1 $M = 20$ (patrz rysunek 4(c)) niż jego $M = 160$ (patrz rysunek 4(d)).

Te wyniki eksperymentalne sugerują, że wraz ze wzrostem liczby zer w CPT, a w konsekwencji wzrostem złożoności przestrzeni poszukiwań, przydatne jest zwiększenie φ w celu zwiększenia stopnia eksploracji GA. Ta zależność nie ma zastosowania do *Alarm*0.25 i *Alarm*0.3. Początkowo było to dla nas nieco zaskakujące, ale teraz stawiamy hipotezę, że płaskowyzę z kondycją $f = 0$ zaczynają zyskiwać na znaczeniu w procesie wyszukiwania. Ponieważ te płaskowyzę są eksplorowane w ten sam sposób przez Generalized Crowding, niezależnie od wartości φ (prawdopodobieństwo wyboru dziecka wynosi 0,5), wydajność GA dla różnych wartości φ jest coraz bardziej podobna. Istnieją sposoby na bardziej efektywną eksplorację tych płaskowyzów, co wykracza poza zakres tego artykułu.

Kluczowe punkty tych eksperymentów są następujące:

- Fakt, że Generalized Crowding pozwala na zmianę selektywnej presji za pomocą parametru φ jest ważny, ponieważ optymalna wartość φ zależy od tego, czy przetwarzany jest *Alarm*0.0, *Alarm*0.1, czy *Alarm*0.2 i która wartość dla M jest używana.
- Deterministyczny Crowding ($\varphi = 0$) i Probabilistyczny Crowding ($\varphi = 1$) generalnie osiągają dobre wyniki. Co zaskakujące, w niektórych przypadkach Generalized Crowding z $\varphi > 1$ daje lepsze wyniki. Na przykład pokazuje to *Alarm*0.2, gdy rozmiar populacji jest zmniejszony - patrz rysunek 4(e) i w mniejszym stopniu rysunek 4(c).

³Ze względu na ograniczone miejsce, szczegółowo przedstawiamy tylko podzbiór wyników eksperymentalnych, ale podsumowujemy szerszy zakres eksperymentów w tekście.



Rysunek 4: Wpływ zmiany współczynnika skalowania (sf lub ϕ) podczas szacowania najbardziej prawdopodobnego wyjaśnienia (MPE) dla różnych wariantów Alarmu BN (Alarm0.0, Alarm0.1 i Alarm0.2) dla dwóch różnych wielkości populacji $M = 20$ i $M = 160$.

Chociaż utrzymywaliśmy stałą wartość ϕ podczas wyszukiwania, można ją dynamicznie dostosowywać. Istnieją różne sposoby, aby to zrobić, co będzie przedmiotem przyszłych badań, ale można (i) wykonywać liniowe dekrementacje przez pokolenia lub (ii) dekrementować proporcjonalnie do entropii populacji.

5.2 Różne sieci bayesowskie

W tej sekcji omówimy wyniki eksperymentów z wykorzystaniem kilku różnych BN, a mianowicie Barley, HeparII i Water. Sieć Barley BN ma 48 węzłów z 130180 wartościami CPT (żadna z nich nie jest równa 0).⁴ HeparII BN zawiera 70 węzłów z 2139 wartościami CPT (żadna z nich nie jest równa 0).⁵ Water BN ma 32 węzły z 13484 wartościami CPT (3972 z nich równych 0).⁶ Wyniki eksperymentalne są uśrednione dla 350 przebiegów, z wyjątkiem jęczmienia, dla którego wykorzystano 50 przebiegów.

Eksperymenty z tymi trzema BN zostały podsumowane na rys.

⁴Szczegółowe informacje na temat jęczmienia można znaleźć na stronie <http://compbio.cs.huji.ac.il/Repository/Datasets/barley/barley.htm>.

⁵Zobacz

<http://www.pitt.edu/~druzdzel/abstracts/springer00-.html> w celu uzyskania szczegółowych informacji na temat HeparII.

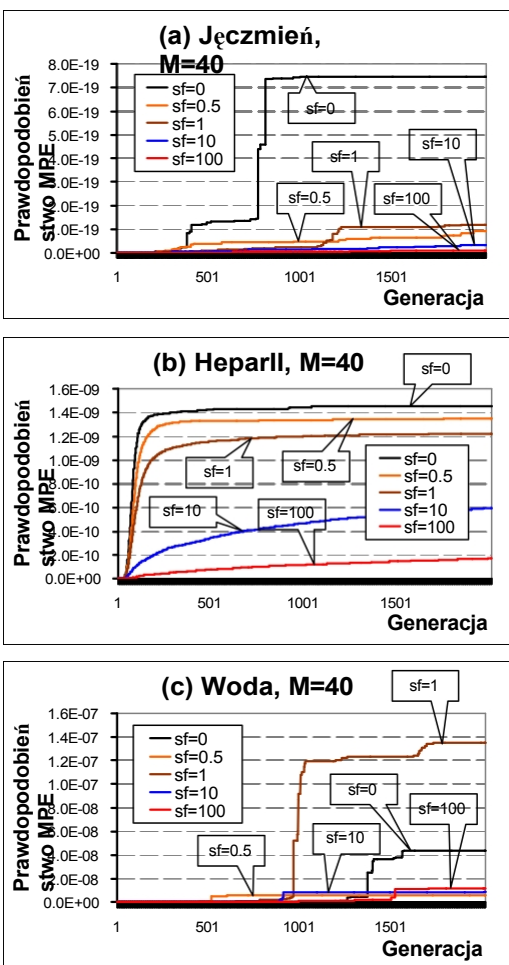
⁶Szczegółowe informacje na temat wody można znaleźć na stronie <http://compbio.cs.huji.ac.il/Repository/Datasets/water/water.htm>.

ure 5, z którego wynika kilka spostrzeżeń. (i) W przypadku BN HeparII i Barley jasne jest, że niższy φ prowadzi do lepszej wydajności. Te dwa BN nie mają zer w swoich CPT. (ii) W przypadku BN Woda (29,45% zer w CPT), $\varphi = 1$ (Probabilistic Crowding) jest lepszy od pozostałych. Oba

(i) i (ii) są zgodne z wynikami uzyskanymi dla *Alarm* ρ : Gdy nie ma zer w CPT, najlepszą opcją jest Determin- istic Crowding; jednakże, gdy zera są obecne w CPT, $\varphi > 0$ prowadzi do najlepszej wydajności. Optymalna *wartość* φ zależy od BN.

6. WNIOSKI I PRZYSZŁE PRACE

W niniejszej pracy zdefiniowano, przeanalizowano i oceniono ogólną metodę tłumienia w GA. Metoda ta opiera się na uelastycznieniu fazy zastępowania w crowdingu, w szczególności poprzez wprowadzenie współczynnika skalowania φ , który umożliwia łatwe kontrolowanie selektywnej presji. Dzięki tej elastyczności nowa metoda, Generalized Crowding, może być z powodzeniem stosowana w szerokim zakresie problemów. Co więcej, Uogólniony Crowding ma jako specjalne przypadki inne szeroko stosowane metody crowdingu, takie jak Deterministyczny Crowding ($\varphi = 0$) i Probabilistyczny Crowding ($\varphi = 1$).



Rysunek 5: Wyniki estymacji MPE dla różnych współczynników skalowania (ϕ lub φ) dla trzech różnych sieci bayesowskich, wielkość populacji $M = 40$.

Właściwa kontrola współczynnika skalowania ϕ , w zależności od problemu przetwarzanego przez Generalized Crowding, stanowi ważny kierunek przyszłych badań. W niniejszym artykule wykazano analitycznie i empirycznie znaczenie współczynnika ϕ , w tym jak jego optymalna wartość zależy od trudności konkretnego problemu i wielkości populacji. Ponieważ trudność problemu zwykle nie jest znana z góry, techniki adaptacyjne lub samoadaptacyjne mogą być wykorzystane do kontrolowania wartości ϕ w trybie on-line. Drugim kierunkiem przyszłych prac jest wykorzystanie przedstawionych tutaj eksperymentów, w których skupiliśmy się na estymacji MPE. Interesujące byłoby rozważenie oszacowania wielu wysoce prawdopodobnych wyjaśnień, w szczególności w sytuacjach, w których takie wyjaśnienia są rozproszone w przestrzeni wyszukiwania.

7. PODZIĘKOWANIA

Niniejszy materiał jest częściowo oparty na pracach Ole J. Mengshoela wspieranych przez granty NSF CCF-0937044 i ECCS-0931978. Dziękujemy anonimowym recenzentom za ich komentarze, które pomogły ulepszyć artykuł.

8. ODNIESIENIA

- [1] D. H. Ackley, G. E. Hinton i T. J. Sejnowski. A learning algorithm for Boltzmann machines. *Cognitive Science*, 9:147-169, 1985.
- [2] K. A. de Jong. *An Analysis of the Behavior of a Class of Genetic Adaptive Systems*. PhD thesis, Department of Computer and Communication Sciences, University of Michigan, Ann Arbor, MI, 1975.
- [3] S. F. Gal'an i O. J. Mengshoel. Obsługa ograniczeń przy użyciu selekcji turniejowej: Abdukcyjne wnioskowanie w częściowo deterministycznych sieciach bayesowskich. *Evolutionary Computation*, 17(1):55-88, 2009.
- [4] D. E. Goldberg. *Algorytmy genetyczne w wyszukiwaniu, optymalizacji i uczeniu maszynowym*. Addison-Wesley, Reading, MA, 1989.
- [5] D. E. Goldberg. Uwaga na temat selekcji turniejowej Boltzmana dla algorytmów genetycznych i symulowane wyżarzanie zorientowane na populację. *Complex Systems*, 4:445-460, 1990.
- [6] J. H. Holland. *Adaptation in Natural and Artificial Systems*. The University of Michigan Press, Ann Arbor, MI, 1975. Wydanie drugie, The MIT Press, Boston, MA, 1992.
- [7] S. Kirkpatrick, C. D. Gelatt i M. P. Vecchi. Optymalizacja przez symulowane wyżarzanie. *Science*, 220(4598):671-680, 1983.
- [8] S. W. Mahfoud. Crowding and preselection revisited. In R. Manner and B. Manderick, editors, *Proceedings of the 2nd International Conference on Parallel Problem Solving from Nature (PPSN II)*, pages 27-36, Brussels, Belgium, 1992. Elsevier, Amsterdam, Holandia.
- [9] S. W. Mahfoud. *Niching Methods for Genetic Algorithms*. Praca doktorska, Department of General Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, 1995.
- [10] S. W. Mahfoud i D. E. Goldberg. Równoległe rekombinacyjne symulowane wyżarzanie: A genetic algorithm. *Parallel Computing*, 21:1-28, 1995.
- [11] O. J. Mengshoel. *Efficient Bayesian Network Inference: Genetic Algorithms, Stochastic Local Search, and Abstraction*. Praca doktorska, Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana, IL, 1999.
- [12] O. J. Mengshoel i D. E. Goldberg. Probabilistic crowding: Deterministyczny tok z probabilistycznym zastępowaniem. In W. Banzhaf, J. M. Daida, A. E. Eiben, M. H. Garzon, V. Honavar, M. J. Jakiela, and R. E. Smith, redaktorzy, *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-1999)*, strony 409-416, Orlando, FL, 1999. Morgan Kaufmann, San Francisco, CA.
- [13] O. J. Mengshoel i D. E. Goldberg. The crowding approach to niching in genetic algorithms. *Evolutionary Computation*, 16(3):315-354, 2008.
- [14] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller i E. Teller. Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, 21(6):1087-1092, 1953.