Adaptacja macierzy kowariancji w metodach strategii ewolucyjnych

Eryk Warchulski
Opiekun pracy:
dr hab. inż. **Jarosław J. Arabas**, prof. PW

Politechnika Warszawska
Wydział Elektroniki i Technik Informacyjnych

Spis treści

Metaheurystyki

Fakt

Heurystyka jest metodą przeszukiwania pewnej przestrzeni $\mathcal{S}$, która nie gwarantuje znalezienia rozwiązania optymalnego według przyjętego kryterium.

Fakt

Heurystyka jest metodą przeszukiwania pewnej przestrzeni $\mathcal{S}$, która nie gwarantuje znalezienia rozwiązania optymalnego według przyjętego kryterium.

Przykłady przestrzeni $\mathcal{S}$:

Fakt

Heurystyka jest metodą przeszukiwania pewnej przestrzeni $\mathcal{S}$, która nie gwarantuje znalezienia rozwiązania optymalnego według przyjętego kryterium.

Przykłady przestrzeni $\mathcal{S}$:

- $\mathcal{S} = \mathbb{R}^n$

Fakt

Heurystyka jest metodą przeszukiwania pewnej przestrzeni $\mathcal{S}$, która nie gwarantuje znalezienia rozwiązania optymalnego według przyjętego kryterium.

Przykłady przestrzeni $\mathcal{S}$:

- $\mathcal{S} = \mathbb{R}^n$
- $\mathcal{S} = \mathbb{G}(V, E)$

Fakt

Heurystyka jest metodą przeszukiwania pewnej przestrzeni $\mathcal{S}$, która nie gwarantuje znalezienia rozwiązania optymalnego według przyjętego kryterium.

Przykłady przestrzeni $\mathcal{S}$:

- $\mathcal{S} = \mathbb{R}^n$
- $\mathcal{S} = \mathbb{G}(V, E)$
- $\mathcal{S} = \Sigma^*$

- na początku XX wieku zaczęto stosować pojęcie *metaheurystyki* bez podania ścisłej definicji

- na początku XX wieku zaczęto stosować pojęcie *metaheurystyki* bez podania ścisłej definicji
- w roku 2006 odbyła się pierwsza konferencja poświęcona metaheurystykom

- na początku XX wieku zaczęto stosować pojęcie *metaheurystyki* bez podania ścisłej definicji
- w roku 2006 odbyła się pierwsza konferencja poświęcona metaheurystykom
- w dziedzinie tej panowała do pewnego momentu *wojna paradygmatów*:

metaphor-centric vs *framework-centric*

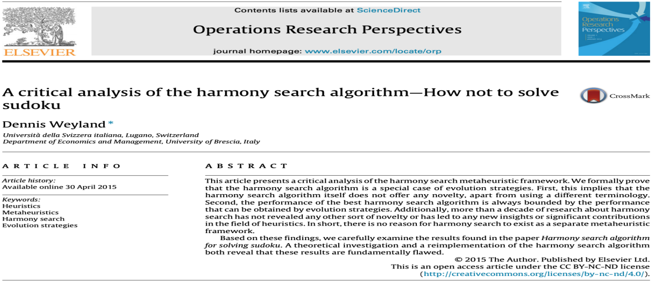
- na początku XX wieku zaczęto stosować pojęcie *metaheurystyki* bez podania ścisłej definicji
- w roku 2006 odbyła się pierwsza konferencja poświęcona metaheurystykom
- w dziedzinie tej panowała do pewnego momentu *wojna paradygmatów*:

metaphor-centric vs *framework-centric*

Eksplozja publikacyjna

Rysunek: Oś czasu z datami publikacji nowych metod metaheurystycznych. Źródło [3].

Akcja-reakcja



Contents lists available at ScienceDirect

Operations Research Perspectives

journal homepage: www.elsevier.com/locate/orp

ELSEVIER

A critical analysis of the harmony search algorithm—How not to solve sudoku

Dennis Weyland *
Università della Svizzera italiana, Lugano, Switzerland
Department of Economics and Management, University of Brescia, Italy

ARTICLE INFO

Article history:
Available online 30 April 2015

Keywords:
Heuristics
Metaheuristics
Harmony search
Evolution strategies

ABSTRACT

This article presents a critical analysis of the harmony search metaheuristic framework. We formally prove that the harmony search algorithm is a special case of evolution strategies. First, this implies that the harmony search algorithm itself does not offer any novelty, apart from using a different terminology. Second, the performance of the best harmony search algorithm is always bounded by the performance that can be obtained by evolution strategies. Additionally, more than a decade of research about harmony search has not revealed any other sort of novelty or has led to any new insights or significant contributions in the field of heuristics. In short, there is no reason for harmony search to exist as a separate metaheuristic framework.

Based on these findings, we carefully examine the results found in the paper *Harmony search algorithm for solving sudoku*. A theoretical investigation and a reimplementation of the harmony search algorithm both reveal that these results are fundamentally flawed.

© 2015 The Author. Published by Elsevier Ltd.
This is an open access article under the CC BY-NC-ND license
(http://creativecommons.org/licenses/by-nc-nd/4.0/).

Rysunek: Fragment krytycznego artykułu wobec metaheurystyk. Źródło [5].

Podejście algebraiczne

Podejście algebraiczne

Definicja

Metaheurystyka jest wysokopoziomowym podejściem do rozwiązywania problemów obliczeniowych.

Podejście algebraiczne

Definicja

Metaheurystyka jest wysokopoziomowym podejściem do rozwiązywania problemów obliczeniowych.

Dla problemu $\mathcal{P}(D, Z)$, w którym D nazywa się dziedziną problemu, a Z przeciwdziedziną, przy pomocy zbioru operacji $\mathcal{O}$ poszukuje się rozwiązania tego problemu.

Podejście algebraiczne

Definicja

Metaheurystyka jest wysokopoziomowym podejściem do rozwiązywania problemów obliczeniowych.

Dla problemu $\mathcal{P}(D, Z)$, w którym D nazywa się dziedziną problemu, a Z przeciwdziedziną, przy pomocy zbioru operacji $\mathcal{O}$ poszukuje się rozwiązania tego problemu.
Innymi słowy – metaheurystyki to sposób myślenia o problemach w oderwaniu od ich reprezentacji.

Obliczenia ewolucyjne

Specyfikacja

$$
\begin{gathered}
\text{typ}: EA(D, Z, M) \\
\text{operacje}: \\
\texttt{init}: Dx\mathbb{N} \rightarrow \text{Zbiór}(Z) \\
\texttt{evol}: POP \rightarrow \text{Zbiór}(Z) \\
\texttt{eval}: POP \rightarrow E \\
\texttt{best}: POP \rightarrow Z \\
\texttt{tune}: M \rightarrow M \\
\texttt{aksjomaty}: z \in POP \implies POP.eval(best()) \leqslant POP.eval(z)
\end{gathered} \tag{1}
$$

- Operacja `init()` inicjalizuje zbiór rozwiązań dla problemu wejściowego $d \in D$, tj. **populację** `POP`, o pewnej liczebności $n \in \mathbb{N}$.

- Operacja `init()` inicjalizuje zbiór rozwiązań dla problemu wejściowego $d \in D$, tj. **populację** `POP`, o pewnej liczebności $n \in \mathbb{N}$.
- na zbiorze tym przeprowadzana jest operacja *evol()* reprezentująca ewolucję, tj. iteracyjną procedurę stosowania operatorów ewolucyjnych jak **mutacja** lub **krzyżowanie**.

- Operacja `init()` inicjalizuje zbiór rozwiązań dla problemu wejściowego $d \in D$, tj. **populację** `POP`, o pewnej liczebności $n \in \mathbb{N}$.
- na zbiorze tym przeprowadzana jest operacja *evol()* reprezentująca ewolucję, tj. iteracyjną procedurę stosowania operatorów ewolucyjnych jak **mutacja** lub **krzyżowanie**.
- na populacji przeprowadzona jest ewaluacja jej **osobników**, która przypisuje im **przystosowanie**.

- Operacja `init()` inicjalizuje zbiór rozwiązań dla problemu wejściowego $d \in D$, tj. **populację** `POP`, o pewnej liczebności $n \in \mathbb{N}$.
- na zbiorze tym przeprowadzana jest operacja *evol()* reprezentująca ewolucję, tj. iteracyjną procedurę stosowania operatorów ewolucyjnych jak **mutacja** lub **krzyżowanie**.
- na populacji przeprowadzona jest ewaluacja jej **osobników**, która przypisuje im **przystosowanie**.
- algorytm ewolucyjny jest ponadto wzbogacony o zbiór parametrów M charakteryzujących jego przebieg (np. wielkość populacji), który może ulegać zmianie w trakcie działania algorytmu.

- algorytm genetyczny Holland'a

- algorytm genetyczny Holland'a
- programowanie genetyczne Kozy

- algorytm genetyczny Holland'a
- programowanie genetyczne Kozy
- ewolucja różnicowa Storn'a

- algorytm genetyczny Holland'a
- programowanie genetyczne Kozy
- ewolucja różnicowa Storn'a
- programowanie ewolucyjne Fogel'a juniora

- algorytm genetyczny Holland'a
- programowanie genetyczne Kozy
- ewolucja różnicowa Storn'a
- programowanie ewolucyjne Fogel'a juniora
- neuroewolucja Fogel'a seniora

Intriguing properties of neural networks

Christian Szegedy Google Inc.
Wojciech Zaremba New York University
Ilya Sutskever Google Inc.
Joan Bruna New York University
Dumitru Erhan Google Inc.
Ian Goodfellow University of Montreal
Rob Fergus New York University, Facebook Inc.

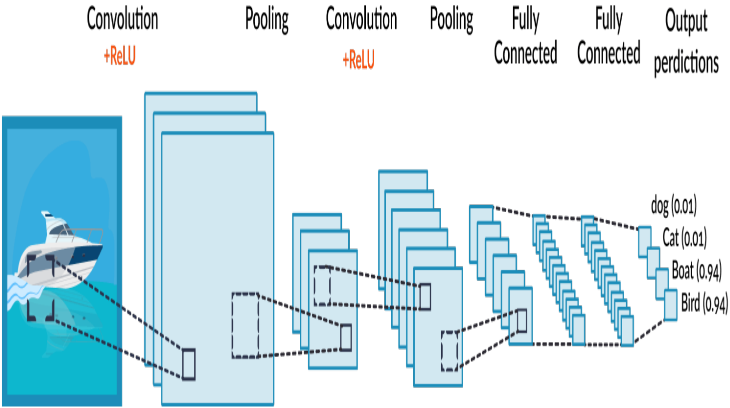
Abstract

Deep neural networks are highly expressive models that have recently achieved state of the art performance on speech and visual recognition tasks. While their expressiveness is the reason they succeed, it also causes them to learn uninterpretable solutions that could have counter-intuitive properties. In this paper we report two such properties.

First, we find that there is no distinction between individual high level units and random linear combinations of high level units, according to various methods of unit analysis. It suggests that it is the space, rather than the individual units, that contains the semantic information in the high layers of neural networks.

Second, we find that deep neural networks learn input-output mappings that are fairly discontinuous to a significant extent. We can cause the network to misclassify an image by applying a certain hardly perceptible perturbation, which is found by maximizing the network's prediction error. In addition, the specific nature of these perturbations is not a random artifact of learning: the same perturbation can cause a different network, that was trained on a different subset of the dataset, to misclassify the same input.

Rysunek: Fragment artykułu dotyczącego zaburzania konwolucyjnych sieci neuronowych. Źródło [4].



Rysunek: Architektura konwolucyjnej sieci neuronowej. Źródło: `Google Images`.

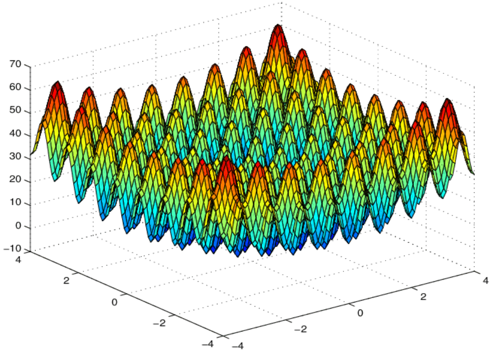
W stronę CMA

- domena metody: $D = \mathbb{R}^{\mathbb{R}^p}$

- domena metody: $D = \mathbb{R}^{\mathbb{R}^p}$
- rezygnacja z krzyżowania $\texttt{evol} = \texttt{mut()}$

- domena metody: $D = \mathbb{R}^{\mathbb{R}^p}$
- rezygnacja z krzyżowania $\texttt{evol} = \texttt{mut()}$
- wyposażenie w samoadaptację $\texttt{tune}: POPxM \rightarrow M$

- domena metody: $D = \mathbb{R}^{\mathbb{R}^p}$
- rezygnacja z krzyżowania $\texttt{evol} = \texttt{mut()}$
- wyposażenie w samoadaptację $\texttt{tune}: POPxM \rightarrow M$
- bazowanie na rozkładzie normalnym $\mathcal{N}_p(\mu, \sigma^2\Sigma)$



Rysunek: Przykład funkcji Rastrigina $f\colon \mathbb{R}^2 \times \mathbb{R}$. Źródło: `Google Images`.

Dlaczego rozkład normalny?

$$x_k := m_k + \sigma_k \mathcal{N}_p(0, \Sigma), k \in \{1, \ldots, \lambda\} \tag{2}$$

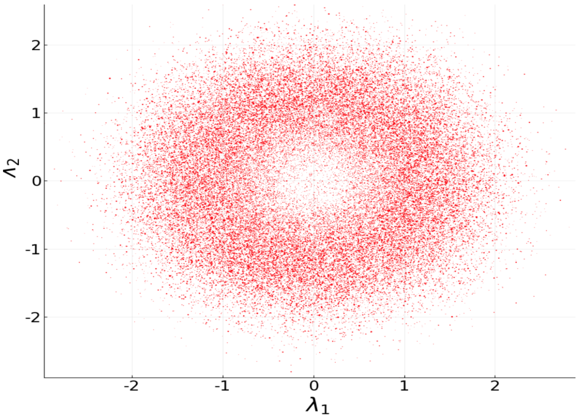
- algorytm w początkowych iteracjach potrzebuje dużej różnorodości, aby móc sprawnie eksplorować przestrzeń przeszukiwań

- algorytm w początkowych iteracjach potrzebuje dużej różnorodości, aby móc sprawnie eksplorować przestrzeń przeszukiwań
- niewielka różnorodność zmniejsza prawdopodobieństwo znalezienia optimum globalnego

- algorytm w początkowych iteracjach potrzebuje dużej różnorodości, aby móc sprawnie eksplorować przestrzeń przeszukiwań
- niewielka różnorodność zmniejsza prawdopodobieństwo znalezienia optimum globalnego
- ...rozkład normalny posiada największą entropię w klasie rozkładów ciągłych z nośnikiem R^p

- algorytm w początkowych iteracjach potrzebuje dużej różnorodości, aby móc sprawnie eksplorować przestrzeń przeszukiwań
- niewielka różnorodność zmniejsza prawdopodobieństwo znalezienia optimum globalnego
- ...rozkład normalny posiada największą entropię w klasie rozkładów ciągłych z nośnikiem R^p

$$-\int_X \mathrm{pdf}(x) \log\{\mathrm{pdf}(x)\} dx \tag{3}$$



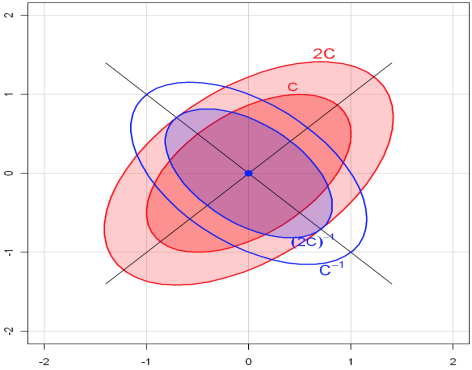
Rysunek: Punkty wylosowane z izotropicznego rozkładu normalnego.`Google Images`.

- kształt rozkładu normalnego jest determinowany przez Σ

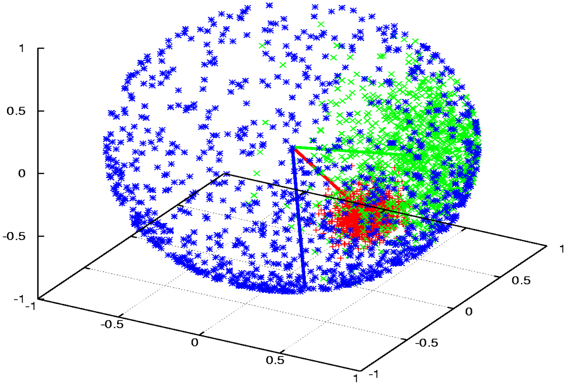
- kształt rozkładu normalnego jest determinowany przez Σ

Elipsoida koncentracji:

$$\{x \in \mathbb{R}^p : (x-\mu)^T \Sigma^{-1} (x-\mu) = c\} \tag{4}$$



Rysunek: Elipsoida kocentracji nie-izotropicznej macierzy kowariancji.`Google Images`.



Rysunek: Niesymetryczny rozkład kierunkowy Fishera-Misesa.`Google Images`.

Samoadaptacja metody

- mechanizm metody, który ciężko wyprowadzić z podstaw statystycznych, które za nim stoją

Samoadaptacja metody

- mechanizm metody, który ciężko wyprowadzić z podstaw statystycznych, które za nim stoją
- wziąż oparty na heurystykach

Prostsze heurystyki:

Prostsze heurystyki:

- reguła $\frac{1}{5}$

Prostsze heurystyki:

- reguła $\frac{1}{5}$
- meta-ewolucja

czy bardziej złożone jak **CSA** (ang. *cumulative step-size adaptation*).

CSA

Definicja: Ścieżka ewolucyjna

Ścieżką ewolucyjną nazywa się j-wymiarowy wektor:

$$\mathbf{m}^t - \mathbf{m}^{t-j} \tag{5}$$

CSA

Definicja: Ścieżka ewolucyjna

Ścieżką ewolucyjną nazywa się j-wymiarowy wektor:

$$\mathbf{m}^t - \mathbf{m}^{t-j} \tag{5}$$

Definicja: Ścieżka przeszukiwań

Ścieżką przeszukiwań nazywa się wielkość:

$$\mathbf{p}_\sigma^{t+1} = (C^t)^{-\frac{1}{2}} \frac{\mathbf{m}_{t+1} - \mathbf{m}_t}{\sigma^t} + (1-\gamma)\mathbf{p}_\sigma^t \tag{6}$$

CSA c.d

Reguła aktualizacji:

$$\sigma^{t+1} = \sigma^{t} exp\{\alpha \frac{\|\mathbf{p}_{\sigma}^{t}\|}{\mathcal{E}\|\mathcal{N}_{p}(0, \mathcal{I}_{p})\|} - 1\} \tag{7}$$

Estymacja macierzy kowariancji

$$\Sigma^{t+1} = \frac{1}{\mu} \sum_{k=1}^{\mu} (\mathbf{x}_k^t - \mathbf{m}_k^t)(\mathbf{x}_k^t - \mathbf{m}_k^t)^T \tag{8}$$

Heurystyk ciąg dalszy

- CMA-ES w wersji podstawowej wprowadza dwie dodatkowe reguły adaptacji macierzy kowariancji

Heurystyk ciąg dalszy

- CMA-ES w wersji podstawowej wprowadza dwie dodatkowe reguły adaptacji macierzy kowariancji
  - użycie poprzedniej wartość macierzy: Σ^t

$$\Sigma^{t+1} = \frac{1}{\mu}\sum_{k=1}^{\mu}(\mathbf{x}_k^t - \mathbf{m}_k^t)(\mathbf{x}_k^t - \mathbf{m}_k^t)^T + (1-\alpha)\Sigma^t \qquad (9)$$

Heurystyk ciąg dalszy

- CMA-ES w wersji podstawowej wprowadza dwie dodatkowe reguły adaptacji macierzy kowariancji
  - użycie poprzedniej wartość macierzy: Σ^t

$$\Sigma^{t+1} = \frac{1}{\mu}\sum_{k=1}^{\mu}(\mathbf{x}_k^t - \mathbf{m}_k^t)(\mathbf{x}_k^t - \mathbf{m}_k^t)^T + (1-\alpha)\Sigma^t \tag{9}$$

  - użycie ścieżki ewolucyjnej w celu kompensacji straty informacji:

$$\mathbf{p}_\Sigma^t(\mathbf{p}_\Sigma^t)^T \tag{10}$$

Finalnie:

Heurystyk ciąg dalszy

- CMA-ES w wersji podstawowej wprowadza dwie dodatkowe reguły adaptacji macierzy kowariancji
  - użycie poprzedniej wartość macierzy: Σ^t

$$\Sigma^{t+1} = \frac{1}{\mu}\sum_{k=1}^{\mu}(\mathbf{x}_k^t - \mathbf{m}_k^t)(\mathbf{x}_k^t - \mathbf{m}_k^t)^T + (1-\alpha)\Sigma^t \tag{9}$$

  - użycie ścieżki ewolucyjnej w celu kompensacji straty informacji:

$$\mathbf{p}_\Sigma^t(\mathbf{p}_\Sigma^t)^T \tag{10}$$

Finalnie:

$$\Sigma^{t+1} = \frac{1}{\mu}\sum_{k=1}^{\mu}(\mathbf{x}_k^t - \mathbf{m}_k^t)(\mathbf{x}_k^t - \mathbf{m}_k^t)^T + (1-\alpha)\Sigma^t + \beta\mathbf{p}_\Sigma^t\mathbf{p}_\Sigma^t \tag{11}$$

Dalsze kroki i otwarte kwestie

- wyznaczenie *lepszych* estymatorów macierzy kowariancji i średniej

$$\Sigma = \delta S + (1-\delta)F \tag{12}$$

- wyznaczenie *lepszych* estymatorów macierzy kowariancji i średniej

$$\Sigma = \delta S + (1 - \delta)F \tag{12}$$

- zbadanie wpływu rozkładów kierunkowych na zachowanie się algorytmu

Bibliografia

1. "Principled Design of Continuous Stochastic Search: From Theory to Practice", Anne Auger, Nikolaus Hansen
2. "An Algebraic Approach to Population-Based Evolutionary Algorithm Generation", Yujun Zheng
3. "A History of Metaheuristics", Kenneth Sorensen et al.
4. "Intriguing properties of neural networks", Ian Goodfellow et al.
5. "A critical analysis of the harmony search algorithm. How not to solve sudoku", Dennis Weyland