

Evolucijski algoritmi

novе smjernice, novi izazovi

Marin Golub

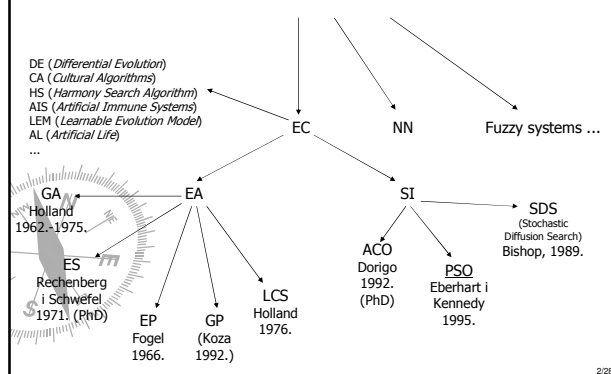
svibanj, 2008.



Gdje su granice evolucijskom računanju?

[A.Menon, editor, *Frontiers of Evolutionary Computation*, Kluwer Academic Publishers, 2004.]

Umjetna inteligencija (AI) Računalna inteligencija (*Computational intelligence - CI*)



Računarska znanost (CS)

- ▶ 1 *Mathematical foundations*
- ▶ 2 *Theory of computation*
- ▶ 3 *Algorithms and data structures*
- ▶ 4 *Programming languages*
- ▶ 5 *Concurrent, parallel, and distributed systems*
- ▶ 6 *Software engineering*
- ▶ 7 *System architecture*
- ▶ 8 *Communications*
- ▶ 9 *Databases*
- ▶ 10 *Artificial intelligence*
- ▶ 11 *Computer graphics*
- ▶ 12 *Human-Computer Interaction*
- ▶ 13 *Scientific computing*
- ▶ 14 *Didactics of computer science/informatics*

Zajedničke osobine evolucijskih algoritama

- ▶ Algoritmi su zasnovani na populaciji rješenja.
- ▶ Jedinke su međusobno usporedive prema dobnosti.
- ▶ Populacija jedinki se s vremenom mijenja, evoluiraju.
- ▶ Svojstva jedinki se prenose s roditelja na djecu.
- ▶ Prostor rješenja se pretražuje slučajnim procesom mutacije.

Poznate činjenice

- ▶ EA s ugrađenim elitizmom asimptotski teži ka globalnom optimumu.
- ▶ Selekcija koja eliminira samo najgore jedinke, odnosno veliki selekcijski pritisak ima za posljedicu ranu konvergenciju ka lokalnom optimumu.
- ▶ EA s premalim selekcijskim pritiskom svodi se na slučajno pretraživanje prostora rješenja.
- ▶ EA nije pogodan za rješavanje deceptivnih optimizacijskih problema.
- ▶ Efikasan evolucijski algoritam ima uravnoteženo slučajno i usmjereno pretraživanje prostora rješenja i uravnoteženu selekciju roditelja i selekciju jedinki za novu populaciju. Kako postići tu ravnotežu?
 - "optimal allocation of trials for SGA" (Holland, 1975)
 - "1/5" rule for ESs (Schwefel, 1981)

Manje poznate činjenice

Selfish gene concept – sve informacije o nekom organizmu zapisane su isključivo u njegovim genima.

Jesu li jedino geni ti koji prijenose svojstva s roditelja na djecu?

Darwin je 1872. u svojem zadnjem izdanju knjige *The Origin of Species by Means of Natural Selection* napisao:

"I am convinced that natural selection has been the main but not the exclusive means of modification."

"... the theory correctly predicts further changes for a short time. But can it explain evolution for a long time? Here the crucial question is: How could it come to such a diversification, starting from a tiny cell?"

Neka neriješena pitanja

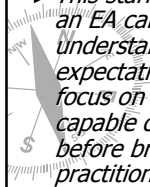
- ▶ *developmental system theory* (Oyama, 2000)
 - EA simuliraju samo prirodni evolucijski proces ne uzimajući u obzir postojanje, djelovanje i razvoj.
 - *Can we formulate a theory of organisms, which incorporates being, acting, evolving, and developing?*
- ▶ Koji su to točno problemi EA teški?
 - deceptijski (a koji su to?)



7/28

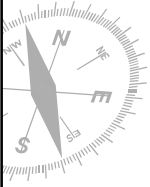
EA su zanimljivi istraživačima

- ▶ Vrlo je jednostavno osmisлити i programski ostvariti EA.
- ▶ Neobično je teško razumijeti zašto on radi. Čak i najjednostavniji SGA.
- ▶ *This stark contrast between the ease with which an EA can be written and the complexity of understanding its behavior leads to a very large expectation gap between EC "practitioners", who focus on the empirical aspects, and often seem capable of thinking of five new genetic operators before breakfast, and the theorists who, to the practitioners, seem fixated on no selection or "counting ones".*



8/28

Smjernice za istraživanja



1. Teorijske osnove

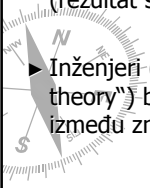
- ▶ Što se očekuje od teorijskih osnova?
 - S 95% vjerojatnošću očekujem da ću EA algoritmom prilikom rješavanja TSP problema s 200 gradova dobiti optimalno rješenje u 10000 iteracija, ako koristim 3-turnirsku selekciju, uniformno križanje i jednostavnu mutaciju s vjerojatnošću 1%.
 - ▶ NE i vjerovatno to nećemo nikada moći s pouzdanošću reći (bez eksperimentalne pripreme)! Previše detalja!
 - Izvesti opće zaključke poput "No Free Lunch" teorema.
 - ▶ NE! Previše općenito!



10/28

Teoretičari i inženjeri

- ▶ Teoretičari („theory in science“ or “scientific theory“) bi trebali razumijeti i pojasniti fenomen (rezultat strogo kontroliranog eksperimenta).
- ▶ Inženjeri („theory in engineering“ or “engineering theory“) bi trebali sagraditi „bolje mostove“ između znanstvenih spoznaja i praktične primjene.



11/28

Zabilježen minoran napredak

- ▶ Jednostavni genetski algoritam (SGA)
 - Teorem šeme (Holland, 1975) i hipoteza građevnih blokova (Goldberg, 1989c; Grefenstette, 1993). To ne vrijedi općenito! Nadalje, znanstvenici su podijeljeni u mišljenju vrijede li navedene teorijske osnove za genetski algoritam općenito ili ne!
 - "messy GAs" [Goldberg, David E., Deb, Kalyanmoy, and Korb, B. (1991). Don't worry, be messy. *Proceedings of the Fourth International Conference on Genetic Algorithms*, pages 24–30. Morgan Kaufmann]
 - [Vose, M. D. (1999). *The simple genetic algorithm: Foundations and theory*. MIT Press, Cambridge, MA]



12/28

Smjernice teorijskih istraživanja

- ▶ modeliranje Markovljevim lancem: Vose model [A.E.Nix, M. D. Vose, 1992. Modeling genetic algorithms with Markov chains, *Annals of Mathematics and Artificial Intelligence*.]
- ▶ *statistical mechanics approach* (Prügel-Bennett and Shapiro, 1994)
- ▶ vjerojatnosti selekcije (seleksijski pritisak i sl.)
- ▶ Postoji li model koji objedinjava teorijska razmatranja za sve evolucijske algoritme?
Sto je zajedničko svim EA?
 - prikaz rješenja (*genotype-phenotype map*)
 - funkcija dobrote
 - skup (evolucijskih?) operatora

13/28

2. EA i slučajno pretraživanje

- ▶ Koliko je (ako je) neki EA bolji od slučajnog pretraživanja ili nekih drugih Monte Carlo metoda?
- ▶ Sustavno ispitivanje na nizu optimizacijskih problema?
 - Simulirano kaljenje
Simulated Annealing, SA (Kirkpatrick, 1983)
 - *Tabu Search, TS* (Glover and Laguna, 1997)
 - ...

14/28

3. Radno okruženje za evolucijsko računanje

- ▶ Razvoj radnog okruženja za evolucijsko računanje.
- ▶ „*developing a unifying EC framework is a key requirement for continued growth and development of the field*”
- ▶ poligon za eksperimentiranje
 - razni EC algoritmi
 - optimizacijski problemi
 - razni pristupi
 - ▶ prikaz, dobrota,
 - ▶ operatori, parametri, ...

15/28

Poligon za eksperimentiranje

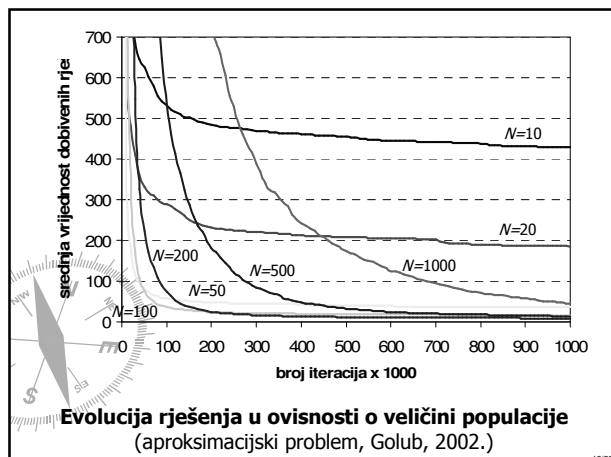
- ▶ *Quadratic Assignment Problems*
- ▶ *Capital Budgeting Problems*
- ▶ *Multiple / Quadratic Knapsack Problems*
- ▶ *Task Allocation Problems (distributed computer systems)*
- ▶ *Maximum Diversity Problems*
- ▶ *P-Median Problems*
- ▶ *Asymmetric and Symmetric Assignment Problems*
- ▶ *Side Constrained Assignment Problems*
- ▶ *Constraint Satisfaction Problems (CSPs)*
- ▶ *Set Partitioning Problems*
- ▶ *Fixed Charge Warehouse Location Problems*
- ▶ *Maximum Clique Problems*
- ▶ *Maximum Independent Set Problems*
- ▶ *Maximum Cut Problems*
- ▶ *Graph Coloring Problems*
- ▶ *Graph Partitioning Problems*
- ▶ problem rasporeda, trgovačkog putnika, aproksimacijski problem, ...

16/28

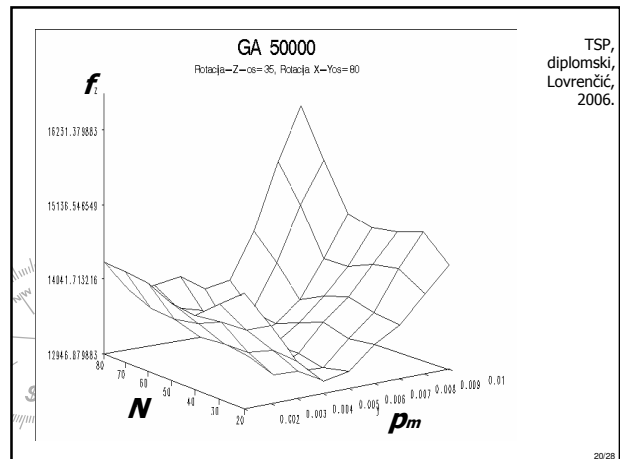
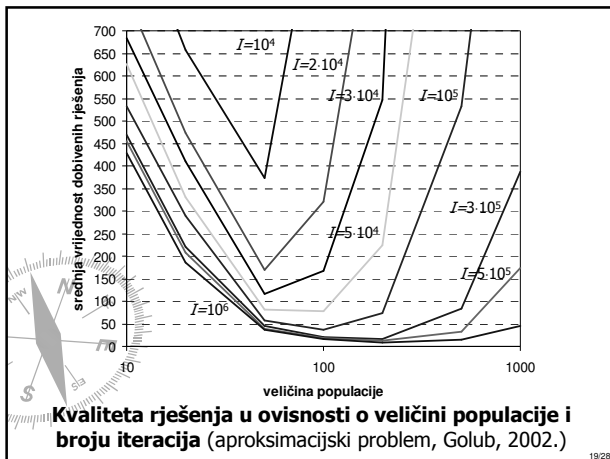
4. Problem izbora optimalnih genetskih operatora i optimalnog skupa parametara

- ▶ Postoje li pravilnosti za neki EA pripravljen za rješavanje određenog optimizacijskog problema?
- ▶ Vrijede li pravilnosti općenito?
- ▶ Reducirani problem: koji je optimalni skup parametara?
- ▶ krajolik parametara (*parameter landscape*)

17/28



18/28







- ### 5. PGA
- ▶ još jedan nepresušan izvor tema...
 - ▶ osnovni modeli su poznati:
 - DGA,
 - GPGA,
 - MPGA.
 - ▶ Eksperimentalna provjera Darwinove pretpostavke ciklusa kontinent-otok.
 - ▶ DGA

- ### 6. EA i dinamički problemi optimiranja
- ▶ Je li EA pogodan za optimizacijske probleme koji se s vremenom (s brojem iteracija, dinamički) mijenjaju?
 - ▶ "What if fitness landscape varies over time?"
 - ▶ kontradiktorni rezultati
 - eksperimenti s GA i aproksimacijskim problemom su pokazali očekivane izvrsne rezultate
 - u literaturi prevladava sumnja

- ▶ Je li AGA rješenje?
- ▶ AGA / AEA
 - nepresušan izvor tema
 - upitni rezultati

- ### 7. Primjena EA u bioinformatiči
- ▶ Kako i koje informacije su pohranjene u DNK?
 - ▶ O kojoj je količini podataka riječ?
 - ▶ Koja je mjera za količinu podataka u DNK?
 - *C-value* - količina *haploidne* DNK
 - *C* dolazi od "*characteristic*", "*content*" ili "*constant*"
 - $1\text{pg} = 10^{-12}\text{g} \approx 978\text{ M dušičnih baza} \approx 244.5\text{ MB}$
 - *junk (or noncoding) DNA* – dio DNK koji se ne koristi i za koji se još ne zna funkcija
 - 80-90% ljudskog genoma je *beskoristan*
 - Paradoks količine DNK (*C-value enigma or C-value paradox*)

Kapacitet genoma (haploidne DNK)

- ▶ čovjek 3 pg \equiv 750Mb 
- ▶ šišmiš (*Miniopterus schreibersi*) 1.7 pg* 
- ▶ miš (*Micoureus demerarae*) 4.9 pg* 
- ▶ *Mesostoma ehrenbergii* 16 pg* (!?)  **

*||| Izvor: [Gregory, T.R. (2008). *Animal Genome Size Database*. <http://www.genomesize.com>]
** simbolična fotografija

25/28

Primjeri primjene EA u bioinformatički

- ▶ GA u potrazi za istim DNK sekvencama u različitim genomima (čovjeka, miša, svinje, kokoši i ribe)
 - [M.A.Lones at all, Regulatory Motif Discovery Using a Population Clustering EA, IEEE/ACM trans. on Computational Biology and Bioinformatics, jul-sep 2007.]
 - [C.B.Congdon at all, An Evaluating of Information Content as a Metric for the Inference of Putative Conserved Noncoding Regions in DNA Sequences Using a GA Approach, IEEE/ACM trans. on Computational Biology and Bioinformatics, jan-mar 2008.]
- ▶ predikcija strukture RNK
 - [K.C.Wiese at all, RnaPredict-An EA for RNA Secondary Structure Prediction, IEEE/ACM trans. on Computational Biology and Bioinformatics, jan-mar 2008.]

26/28

Primjeri primjene EA u bioinformatički

- ▶ GA za optimalno grupiranje aminokiselina
 - [J.Lenckowski, K.Walczak Simplifying Amino Acid Alphabets Using a Genetic Algorithm and Sequence Alignment, LNCS 4447, EvoBIO 2007.]
- ▶ EA i DE za optimiranje procesa fermentacije
 - [M.Rocha at all, Evaluating Evolutionary Algorithms and Differential Evolution for the Online Optimization of Fermentation Processes, LNCS 4447, EvoBIO 2007.]
- ▶ doziranje lijekova
 - [F. Archetti at all, Genetic Programming and Other Machine Learning Approaches to Predict Median Oral Lethal Dose (LD50) and Plasma Protein Binding Levels (%PPB) of Drugs, LNCS 4447, EvoBIO 2007.]

27/28

Umjesto zaključka

- ▶ *uzbudljivo* područje istraživanja
- ▶ tema ima na pretek
 - ima mnogo posla ☺
- ▶ iz godine u godinu bilježi se *superlinearan* broj istraživača i broj objavljenih radova...

materijali su dostupni na
<http://www.zemris.fer.hr/~golub/ga/ga.html>

28/28