Metodologia cercetarii in Informatica

Catalin Stoean
catalin.stoean@inf.ucv.ro
http://inf.ucv.ro/~cstoean

Cuprinsul cursului

1. Ce este cercetarea?
2. Alegerea unui subiect de cercetare
3. Cum citim un articol
4. Formularea problemelor
5. Evaluare si validare
6. Publicatii la conferinte si reviste
7. Scrierea de lucrari tehnice (incl. disertatie)
8. Prezentarea lucrarilor
9. Scrierea de propuneri de grant
10. Etica

Conferinte si Reviste

- Reputatie editori
- Asociatii profesionale recunoscute: IEEE, ACM etc.
- Stati departe de forumuri comerciale cu scopul de a face bani din taxe: WSEAS, IAENG, DAAAM
  - http://fakeconference.blogspot.ro/
  - http://www.cs.bris.ac.uk/Teaching/learning/junk.conferences.html
  - http://en.wikipedia.org/wiki/World_Multiconference_on_Systemics,_Cybernetics_and_Informatics

Calitate si metrici

- Factorul de impact
  - Reflecta numarul mediu de citari la articolele publicate recent in revista.
  - Releva importanta revistei in domeniul sau.
  - Se calculeaza ca numarul mediu de citari primit pentru o lucrare in ultimii 2 ani.
- Citarile cuantifica impactul unei lucrari.
- Scorul de influenta al unei reviste
  - Daca este peste 1, fiecare articol din revista are o influenta peste medie.

De ce se organizeaza conferinte?

- Pentru a promova schimbul de idei intr-un domeniu
- Pentru a promova colaborarea intre cercetatori
- Pentru a genera fonduri pentru o organizatie non-profit
- Pentru a atrage atentia asupra unui domeniu
- Pentru a promova reputatia organizatorilor
- Pentru a genera un profit pentru organizatori

De ce se merge la conferinte?

- Pentru a invata despre un domeniu
- Pentru a interactiona cu alti cercetatori
- Pentru a adauga o publicatie la CV
- Pentru a avea o vacanta intr-o locatie exotica

- La ambele intrebari, daca ultimele doua motivatii sunt cele care conteaza se ajunge ca articolele aparute sa fie de calitate slaba.

Calitatea conteaza

- Publicatiile ii ajuta pe cercetatori in cariera:
  - Pentru a atrage fonduri
  - Pentru promovari
  - Pentru a atrage studenti
  - Pentru a fi invitat sa tii prelegeri la alte universitati etc
- Calitatea publicatiilor conteaza insa.
- SCIgen este un software care genereaza articole de cercetare in informatica fara sens.
  - http://en.wikipedia.org/wiki/SCIgen

Calitatea conteaza

- Articole generate de SCIgen au fost acceptate la conferinte:
  - *Rooter: A Methodology for the Typical Unification of Access Points and Redundancy*, World Multiconference on Systemics, Cybernetics and Informatics (WMSCI), 2005.
- Exemplul de mai sus este cel mai celebru, insa si alte conferinte au demonstrat o calitate foarte slaba acceptand astfel de articole.
  - O lucrare[1] din 2013 raporteaza un numar de 85 de lucrari generate de SCIgen au fost publicate chiar de IEEE.
- Reputatia celor care publica la astfel de conferinte are de suferit.

[1]http://hal.archives-ouvertes.fr/docs/00/71/35/55/PDF/0-FakeDetectionSci-Perso.pdf

Cum se recunosc conferintele spam

- Li se face reclama prin spam
- Apelul pentru conferinta subliniaza faptul ca este o conferinta cu reputatie care contine multi experti faimosi
- Subiectul conferintei este unul foarte general
- Conferinta are loc intr-un loc exotic
- Esti invitat de o persoana necunoscuta sa organizezi o sesiune speciala chiar daca nu ai neaparat o pozitie evidenta in domeniul in care activezi.

Revistele si conferintele care conteaza in Informatica

- http://informatica-universitaria.ro/ppages/16/
  - Criteriile de promovare in Informatica
  - Clasificarea jurnalelor din domeniul Informatica
  - Clasificarea conferintelor din domeniul Informatica
- Se pondereaza punctajul dupa numarul de autori.
- Se calculeaza separat ce impact au avut publicatiile pe baza citarilor primite.
- Reputatia internationala conteaza.

Clasificarea revistelor in Informatica

Clasificarea jurnalelor din domeniul Informatică

pozitia	ISSN	Nume jurnal	Categorie
1	03600300	ACM Computing Surveys	A
2	15564665	ACM Transactions on Autonomous and Adaptive Systems	A
3	19423454	ACM Transactions on Computation Theory	A
4	15293785	ACM Transactions on Computational Logic	A
5	07342071	ACM Transactions on Computer Systems	A
6	10730516	ACM Transactions on Computer-Human Interaction	A
7	03625915	ACM Transactions on Database Systems	A
8	07300301	ACM Transactions on Graphics	A
9	10949224	ACM Transactions on Information and System Security	A
10	10468188	ACM Transactions on Information Systems	A
11	15335399	ACM Transactions on Internet Technology	A
12	15564681	ACM Transactions on Knowledge Discovery from Data	A
13	00983500	ACM Transactions on Mathematical Software	A
14	15516857	ACM Transactions on Multimedia Computing, Communications and Applications	A
15	01640925	ACM Transactions on Programming Languages and Systems	A
16	15504859	ACM Transactions on Sensor Networks	A
17	1049331X	ACM Transactions on Software Engineering and Methodology	A
18	15591131	ACM Transactions on the Web	A

Sunt 1326 reviste impartite in 4 categorii, de la A la D.

Productia stiintifica

forumurilor sunt următoarele:

A ➔ 8 pct.

B ➔ 4 pct.

C ➔ 2 pct. Aici se vor include şi LNCS-urile care nu sunt în A şi B.

D ➔ 1 pct. Aici intră reviste şi conferinţe care nu se regăsesc în categoriile de mai sus dar sunt indexate SCOPUS, IEEE, ACM, SPRINGER, DBLP, CiteSeerX, Zentralblatt, MathSciNet etc. (nu şi forumuri de genul WSEAS, IAENG, DAAAM).

Pentru evaluarea perspectivei b) pentru gradele de conferenţiar/CP II şi profesor/CP I doar scorurile asociate lucrărilor ce conţin **rezultate originale ale candidatului** publicate în forumuri de categoriile A, B sau C pot fi însumate. Workshop-urilor asociate conferinţelor obţin rangul conferinţei numai dacă nu sunt cotate separat (caz în care se evaluează ca forumuri independente) şi lucrările lor au fost publicate în volumul conferinţei principale. Lucrările de tip poster şi demonstraţii de sisteme, publicate în volumul conferinţei, precum şi cele prezentate în workshop-uri asociate conferinţelor dar publicate în volume distincte se punctează înjumătăţind scorul conferinţei.

Punctajul publicaţiilor se calculează după formula:

$$P = \sum_{i}^{N} \frac{S_i}{\max(1, n-2)}$$

unde S_i reprezintă scorul asociat forumului unde a apărut publicaţia, iar n este numărul de autori al lucrării.

	Conferenţiar / CP II	**Profesor / CP I**
Valori minimale pentru indicatorul *P*	32 puncte, *din care 16 puncte din lucrări de cel puţin categoria B*	56 puncte, *din care 24 puncte din lucrări de categoria A şi 16 puncte din lucrări de cel puţin categoria B*

Impactul rezultatelor

Impactul rezultatelor (perspectiva c) se evaluează pe baza citărilor (fără autocitări). Citările sunt luate în considerare dacă apar în forumuri de tip A, B, C şi D, dar şi în teze de doctorat, monografii şi cărţi editate, unde vor fi echivalate cu tipul D. Punctajul citărilor se calculează după formula:

$$C = \sum_{i=1}^{N} \frac{\sum_{j=1}^{N_i} S_j^i}{\max(1, n_i - 2)}$$

unde: N – numărul de lucrări citate,
n_i – numărul de autori ai lucrării citate,
N_i – numărul de citări ale lucrării i,
S_j^i – scorul asociat forumului în care apare a j-a lucrare care citează lucrarea i.

Auto-citările (citări în articole ale oricăruia dintre autori) se exclud. Citările se consideră pentru orice lucrări ştiinţifice, inclusiv monografii şi articole apărute în publicaţii din afara listelor precizate, atâta timp cât lucrările respective sunt din domeniul Informatică. În acest caz, în calculul punctajului se consideră scorul S_j^i ca fiind 8/4/2/1, în funcţie de categoria editurii conform clasamentului SENSE[1] (în ultima categorie incluzându-se editurile din categoriile D, E şi nelistate).

	Conferenţiar / CP II	**Profesor / CP I**
Valori minimale pentru indicatorul *C*	48, *din care 12 puncte în forumuri de minim tip B*	120, *din care 40 de puncte în forumuri de minim tip B*

Performanta academica 1/3

Performanţa academică (perspectiva d) se evaluează după indicatorii din tabelul de mai jos
(*n* = numărul de autori):

Cărţi autor/editate şi capitole publicate în edituri de categoria (conform clasamentului SENSE): - cărţi (A) - cărţi (B) sau capitole (A) - cărţi (C) sau capitole (B) - cărţi (D, E şi nelistate) sau capitole (C) - capitole (D, E şi nelistate)	 16 / max(1, *n*-2) 8 / max(1, *n*-2) 4 / max(1, *n*-2) 2 / max(1, *n*-2) 1 / max(1, *n*-2)	pe volum sau capitol
Editor proceedings la conferinţe de tip: - A | B | C | D	(8 | 4 | 2 | 1) / max(1, *n*-2)	pe volum
Publicarea unui curs universitar în format electronic	2	pe curs
Director/editor al unei reviste de tip: - A | B | C | D	 24 | 12 | 6 | 3	pe revistă
Director (coordonator/responsabil) | membru al unui grant/proiect/contract/program de cercetare naţional/internaţional a cărui valoare intrată în instituţie este ≥ 200.000 Euro, ca director | membru 100.000 – 199.999 Euro, ca director | membru 50.000 – 99.999 Euro, ca director | membru <50.000 Euro, ca director | membru	 8 | 4 6 | 3 4 | 2 2 | 1	pe grant/proiect/ contract/program

Performanta academica 2/3

Membru în comitetul ştiinţific (de program) al unor conferinţe, simpozioane, workshop-uri, de tip: - A | B | C | D	4 | 2 | 1 | 0,5	pe eveniment
Organizare evenimente ştiinţifice/şcoli de vară, în calitate de: - director | membru în comitetul de organizare	2 | 1	pe eveniment
Keynote/invited speaker la evenimente/universităţi[2]: - de tip A/top 100 - de tip B/şcoli de vară internaţionale/top 200 - de tip C, şcoli de vară naţionale, conferinţe ale Academiei Române/top 500 - de tip D/evenimente locale/top 1000	8 4 2 1	pe eveniment/conferinţă
Profesor/researcher asociat/visiting la o universitate din: - top 100 - top 200 - top 500 - top 1000	8 * nr. luni 4 * nr. luni 2 * nr. luni nr. luni	pe vizită (pentru vizite scurte cu predare intensivă se pot face echivalări: 1 lună=16 ore de predare) Max. 24 puncte
Consolidarea de echipe de cercetare dovedită prin publicaţii, participări în proiecte, dezvoltarea de instrumente software, resurse şi colecţii de date de largă utilitate	nr. persoane * nr. ani de activitate comună	pe echipe de minimum 3 persoane

Performanta academica 3/3

Membru în comisii de evaluare a tezelor de doctorat la o universitate din top: - 100 | 200 | 500 | 1000	4 | 2 | 1 | 0,5	pe teză
Brevete şi invenţii active (OSIM, ORDA etc.)	8 / max(1, *n*-2)	pe brevet/invenţie
Dezvoltarea de pachete şi instrumente software, dezvoltarea de resurse şi colecţii de date de largă utilitate (probate prin număr de accesări, publicarea pe site-uri open source etc.)	2 * nr. produse / max(1, *n*-2)	pe produs
Poziţii de conducere în organizaţii profesionale: - internaţionale | naţionale	4 | 2	pe organizaţie
Premii şi alte merite (la decizia universităţii sau institutului de cercetare)	max 10% din punctajul criteriului	-

	Conferenţiar / CP II	**Profesor / CP I**
Valori minimale pentru perspectiva d	36	60

Observaţie: Un procent de 10% din punctajul total (116 pct. pentru gradele de conferenţiar şi CP II, respectiv 236 pct. pentru gradele de profesor, CP I) poate fi deplasat între criterii, cu respectarea cerinţelor minimale (în italic în tabelele de la perspectivele b şi c).

Cuprinsul cursului

1. Ce este cercetarea?
2. Alegerea unui subiect de cercetare
3. Cum citim un articol
4. Formularea problemelor
5. Evaluare si validare
6. Publicatii la conferinte si reviste
7. Scrierea de lucrari tehnice (incl. disertatie)
8. Prezentarea lucrarilor
9. Scrierea de propuneri de grant
10. Etica

Exprimarea ideii articolului

- Prezentam problema
- Este o problema interesanta
- Este o problema nerezolvata/netratata suficient
- **Iata ideea proprie**
- Ideea functioneaza
  - Detalii, rezultate, observatii
- Aratam cum se compara rezultatele metodei proprii cu cele obtinute de abordarile altor cercetatori

Scriere articol

- Multe lucrari sunt prost scrise si sunt greu de inteles
  - Idei bune pot ramane astfel neapreciate
- Daca urmam niste indicatii simple, calitatea lucrarilor creste
- Cercetarea voastra va fi citita, apreciata, citata
  - Feedback-ul primit ajuta la imbunatatirea cercetarii si a stilului de scriere
- Citirea de articole va familiarizeaza cu modul in care se scrie un articol

Structura

- **Abstract (4 fraze sau cca 150-200 de cuvinte)**
- Introducere (1 pagina)
- Problema (1 pagina)
- Cercetari similare (1-2 pagini)
- Ideea proprie (2 pagini)
- Detalii rezultate (5 pagini)
- Concluzii si idei de continuare a cercetarii (0.5 pagini)

Abstract-ul (sumarul)

- Cel mai bine este sa fie scris la sfarsit, dupa ce este tot articolul gata.
  - Eu asa procedez.
- Este utilizat de membrii din *Program Committee* sa decida la ce recenzori trimit lucrarea.
- 4 fraze
  1. Problema tratata
  2. Motivatia pentru problema (de ce este interesanta)
  3. Solutii pentru problema
  4. Concluzii

Structura

- Abstract (4 fraze sau cca 150-200 de cuvinte)
- **Introducere (1 pagina)**
- Problema (1 pagina)
- Cercetari similare (1-2 pagini)
- Ideea proprie (2 pagini)
- Detalii rezultate (5 pagini)
- Concluzii si idei de continuare a cercetarii (0.5 pagini)

Introducere (1 pag)

- Se descrie problema tratata, eventual cu un exemplu de instanta
- Se indica sumar contributiile anterioare existente si cele proprii aduse prin aceasta cercetare

Multimodal Optimization by Means of a Topological Species Conservation Algorithm

Catalin Stoean, *Member, IEEE,* Mike Preuss, *Member, IEEE,* Ruxandra Stoean, *Member, IEEE,* and D. Dumitrescu

***Abstract*—Any evolutionary technique for multimodal optimization must answer two crucial questions in order to guarantee some success on a given task: How to most unboundedly distinguish between the different attraction basins and how to most accurately safeguard the consequently discovered solutions. This paper thus aims to present a novel technique that integrates the conservation of the best successive local individuals (as in the species conserving genetic algorithm) with a topological subpopulations separation (as in the multinational genetic algorithm) instead of the common but problematic radius-triggered manner. A special treatment for offspring integration, a more rigorous control on the allowed number and uniqueness of the resulting seeds, and a more efficient fitness evaluations budget management further augment a previously suggested naïve combination of the two algorithms. Experiments have been performed on a series of benchmark test functions, including a problem from engineering design. Comparison is primarily conducted to show the significant performance difference to the naïve combination; also the related radius-dependent conserving algorithm is subsequently addressed. Additionally, three more multimodal evolutionary methods, being either conceptually close, competitive as radius-based strategies, or recent state-of-the-art are also taken into account. We detect a clear advantage of three of the six algorithms that, in the case of our method, probably comes from the proper topological separation into subpopulations according to the existing attraction basins, independent of their locations in the function landscape. Additionally, an investigation of the parameter independence of the method as compared to the radius-compelled algorithms is systematically accomplished.**

***Index Terms*—Evolutionary algorithms, function optimization, landscape detection, multimodal optimization, species conservation.**

I. Introduction

MOST OF THE black-box real-world problems considered to be difficult are multimodal. Hence, any optimization technique applied in this area should be able to discover several solutions, namely located in a number of basins of attraction. This enables decision makers to choose from multiple distinct solutions to a problem and, at the same time, increases confidence to have attained the global optimum. Canonical evolutionary algorithms (EA)—despite usually being population-based—have the property of converging to a population that contains only one solution and small variations of it (genetic drift) [1], [2]. In the best case, the fittest obtained solution represents the global optimum, but it may also happen that it only refers to a local optimum in which the search process is confined. In order to achieve an explorative search, EAs that perform multimodal optimization have to either apply multistart techniques or maintain a high diversity in the population with the purpose of searching within many different locations in parallel. Every multimodal optimization method has to consequently satisfy two partly conflicting tasks: to locate the global optimum out of multiple local peaks and to find a set of several good solutions for variety and insights into the problem space.

There have been several attempts for transforming EAs so that they could deal with multimodal fitness landscapes (e.g., [1], [3]–[11]). However, when tailoring such an EA, there are a number of issues to be tackled: 1) how to divide the population into subpopulations; 2) how to preserve these subpopulations in order to avoid the genetic drift; and 3) how to eventually connect them to the existing optima within the fitness landscape. Most techniques for the detection of multiple attraction basins (niching) form subpopulations by appointing a radius such that all individuals within the same species lie at a distance from each other that is lower than the given threshold (they are highly similar). The value that has to be selected for the radius directly depends on the fitness landscape, i.e., on the problem to be solved, whereas its proper choice is crucial in assuring accurate results. Deb and Goldberg [12] proposed a very precise approximation for this parameter, however, especially for real-world applications, the information on the fitness landscape required by the formula is not available beforehand and, therefore, in such situations,

Introducere (1 pag)

- Se enumera contributiile aduse de catre lucrare
- Intreaga lucrare este conceputa in jurul acestor contributii
- Cititorul este motivat de aceste contributii din introducere sa citeasca mai departe.

triggers flexibility as regards the formation of the species within attraction basins of different sizes. Multiple optima maintenance is conducted through the preservation of several distinct solutions. Each species is concentrated on a seed, which represents the fittest individual of the species. The seeds from all species are copied from one generation to another so that no important regions are lost through selection and variation operators. The species masters are then updated at each cycle, by once more appointing their fittest inner individuals.

The manner of detecting whether two individuals follow different peaks or not was initially proposed in [5] and [13], within the multinational genetic algorithm (MGA), but the complete mechanism proved to be very expensive as regards the number of fitness evaluations necessary to converge to the solution [14]. On the other hand, the idea of species conservation first appeared in [6], however, subpopulation differentiation is powered by a radius.

A first attempt to unite the seed preservation and the fitness landscape inspection through a straight integration was the topological species conservation (TSC) approach in [14]. However, the method presented here (TSC2) is significantly improved as it reconsiders the species management to save precious evaluations and accelerate convergence into the basins. Experimentation finally demonstrates its superiority over the initial naïve combination.

The comparison is conducted on several functions that have at least two variables—in order to observe how the optimal peaks are disposed within the landscape—and up to 20, as most real-world problems are multidimensional. The multimodality conditions range from one optimum (the method must still not fail to perform well in the unimodal case) to many global or local peaks environments. Also, a multimodal problem that bears relationship to a generalized real-world application of engineering design is chosen as a test instance. In order to achieve an objective validation, the results obtained by the novel technique are put against those of two other related and recently proposed multimodal EAs [6], [7], the outcomes of a niching strategy [15], and those of a crowding (thus nonradius-based) approach [11]. To demonstrate the important differences to the preliminary integration in [14], TSC2 is also compared to the original TSC.

Structura

- Abstract (4 fraze sau cca 150-200 de cuvinte)
- Introducere (1 pagina)
- **Problema (1 pagina)**
- Cercetari similare (1-2 pagini)
- Ideea proprie (2 pagini)
- Detalii rezultate (5 pagini)
- Concluzii si idei de continuare a cercetarii (0.5 pagini)

Problema

- Descrierea problemei
  - Un exemplu
  - Figura, schema etc
- De ce este interesanta
  - Cum este intalnita in practica
- Citati lucrari relevate pentru problema

Structura

- Abstract (4 fraze sau cca 150-200 de cuvinte)
- Introducere (1 pagina)
- Problema (1 pagina)
- **Cercetari similare (1-2 pagini)**
- Ideea proprie (2 pagini)
- Detalii rezultate (5 pagini)
- Concluzii si idei de continuare a cercetarii (0.5 pagini)

Cercetari similare (state of the art)

- Nu trebuie sa facem cercetarea altora sa arate rau ca sa arate a noastra bine
- Fiti generosi cu cercetatorii care lucreaza in acelasi domeniu:
  - *In his inspiring research [Han], Hansen introduced …*
- Daca apreciati pozitiv munca celorlalti cercetatori nu scade importanta propriei cercetari

triggers flexibility as regards the formation of the species within attraction basins of different sizes. Multiple optima maintenance is conducted through the preservation of several distinct solutions. Each species is concentrated on a seed, which represents the fittest individual of the species. The seeds from all species are copied from one generation to another so that no important regions are lost through selection and variation operators. The species masters are then updated at each cycle, by once more appointing their fittest inner individuals.

The manner of detecting whether two individuals follow different peaks or not was initially proposed in [5] and [13], within the multinational genetic algorithm (MGA), but the complete mechanism proved to be very expensive as regards the number of fitness evaluations necessary to converge to the solution [14]. On the other hand, the idea of species conservation first appeared in [6], however, subpopulation differentiation is powered by a radius.

A first attempt to unite the seed preservation and the fitness landscape inspection through a straight integration was the topological species conservation (TSC) approach in [14]. However, the method presented here (TSC2) is significantly improved as it reconsiders the species management to save precious evaluations and accelerate convergence into the basins. Experimentation finally demonstrates its superiority over the initial naïve combination.

The comparison is conducted on several functions that have at least two variables—in order to observe how the optimal peaks are disposed within the landscape—and up to 20, as most real-world problems are multidimensional. The multimodality conditions range from one optimum (the method must still not fail to perform well in the unimodal case) to many global or local peaks environments. Also, a multimodal problem that bears relationship to a generalized real-world application of engineering design is chosen as a test instance. In order to achieve an objective validation, the results obtained by the novel technique are put against those of two other related and recently proposed multimodal EAs [6], [7], the outcomes of a niching strategy [15], and those of a crowding (thus nonradius-based) approach [11]. To demonstrate the important differences to the preliminary integration in [14], TSC2 is also compared to the original TSC.

The paper is organized as follows. The next section briefly describes some of the traditional evolutionary approaches for multimodal optimization and several new ones that are relevant from the point of view of the design and objectives of proposed technique. The novel method TSC2 is presented in detail in Section III, also highlighting the differences to TSC. Section IV reports on the experimental results comparing to the algorithms named above, and Section V concludes the paper.

II. Evolutionary Speciation Techniques for Multimodal Optimization

In nature, an ecosystem is usually composed of regions (niches) that exhibit different characteristics and allow the formation and maintenance of different types of species. Commonly, the individuals in a species share similar biological features that allow them to coexist in their niches, capable of interbreeding among themselves, but unable to breed with individuals from different species. Each niche is usually populated by a number of individuals that directly depends on the amount of resources the niche provides.

Analogously, in an artificial system, each niche is related to an optimum of the fitness landscape and the resident species contains, in the best case, only individuals being located in the basin of attraction of that peak. In this respect, niching or speciation methods have been proposed for the simultaneous evolution of subpopulations.

A. Radius-Based

The best known niching method is the sharing approach that was initially introduced by Holland [8] and subsequently improved by Goldberg and Richardson [4]. The population is split into several species by taking into account the similarity between individuals. A sharing function modifies the fitness of an individual to be dependent on the number of potential solutions that exist within the same subpopulation. Within the species conserving genetic algorithm (SCGA) in [6], the fittest individuals that are more distant from each other than a predefined radius are set as seeds of their subpopulations. All other individuals (that are not seeds) are each appointed to belong to the subpopulation of the fittest individual that is found within the given radius. The seeds are conserved from one generation to another in order to avoid the risk of extinction following the application of variation operators and they are updated every generation. The SCGA elitist idea of transferring the seeds of each subpopulation from one generation to another is also adopted in the technique proposed herein.

Dynamic fitness sharing (DFS) is introduced in [7]. The technique uses a radius for separating the population into species, allows for a fixed minimum value (of two individuals) for the size of a subpopulation and has, like in the case of SCGA, a dominating individual called the *species master*. This is considered to be the member of the species that has the highest raw fitness value. Within DFS, the subpopulations are identified in each generation using the distance between individuals, while comparing it to the radius threshold. Fitness sharing is employed to compute the weighted fitness of each individual. A species elitist strategy is employed to ensure the conservation of the most prolific individual in each subpopulation from a generation to the other.

The niching variant of the covariance matrix adaptation-evolution strategy (CMA-ES) of Hansen and Ostermeier [16] was introduced by Shir and Bäck [17]. Using a fixed given radius, the population is split into species by means of a technique named dynamic peak identification, so that a predefined number of q niches is generated. This largely resembles a parallel execution of several independent hillclimbers at different locations, separated by a distance of at least the given radius. On recommendation of the authors of [10] who also provided source code for the method, a niching CMA-ES based on q separate $(1+10)$-CMA-ES is employed. These have been proposed by Igel *et al.* [18], are extremely simple and cope well with populations of only one parent individual. The CMA-ES parameters have been shown to be very robust

Cercetari similare

- Daca sugerati ca o idee va apartine si recenzorul stie ca nu este asa, atunci ori
  - Nu stiti ca o idee este veche

  - Stiai ideea, dar pretindeai ca iti apartine

Structura

- Abstract (4 fraze sau cca 150-200 de cuvinte)
- Introducere (1 pagina)
- Problema (1 pagina)
- Cercetari similare (1-2 pagini)
- **Ideea proprie (2 pagini)**
- **Detalii rezultate (5 pagini)**
- Concluzii si idei de continuare a cercetarii (0.5 pagini)

Ideea si experimente

- Descrie ideea propusa
  - Cu detalii
  - Cu schema
  - Cu algoritm(i)
  - Cu citari ale lucrarilor relevante
- Dati o explicatie generala, intuitiva
  - Urmata de detalii
  - Cititorul poate urma astfel textul

Algorithm 1 Detect-Multimodal Mechanism Between Two Individuals x and y

```
1: i = 1;
2: found = FALSE;
3: while i < number of gradations and not found do
4:    for j = 1 to D do
5:       interior_j = x_j + (y_j − x_j) · gradation_j;
6:    end for
7:    if f(interior) < min[f(x), f(y)] then
8:       found = TRUE;
9:    end if
10:   i = i + 1;
11: end while
12: return found;
```

Although robust, this mechanism makes an algorithm more expensive in terms of the number of fitness evaluations, as observed in MGA and TSC [14]. To counteract its effect, a free individual is checked against the seeds in increasing distance order to minimize the number of calls to the *detect-multimodal* procedure.

An important advantage of this manner of detecting multimodality is that it avoids the existence of several subpopulations assigned to follow a certain optimum, as it happens when the radius-based mechanism of species conservation is used. Instead, it assumes the connection of a subpopulation to only one peak, regardless of the size of the basin of attraction of that optimum.

Conversely, when TSC2 deals with a spiny function, with large increments followed by small decrements before rising again, the currently inflicted upper bound for the number of seeds prevents the entire population from being transformed (blocked) into species masters. This blockage would appear as a result of *detect-multimodal* being in charge of establishing them. But, with this limit, only a small part of the population is chosen as seeds. If other good solutions are subsequently found, each is assigned to the closest existing seed (according to the genotype) and, if fitter than the latter, it becomes the current species seed in the next generation.

2) *Conservation—Is It Necessary?:* In every generation, there are a certain number of species, each having its dominating individual and following a different peak. On the one hand, a weighted mating selection is employed, resulting that the fitness of each individual is divided by the size of the species it belongs to. This gives a greater chance to escape extinction to species that have only few individuals, just like

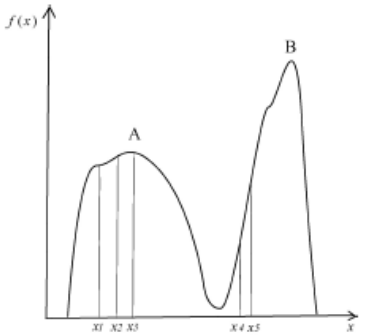

Fig. 1. Valuable individuals could vanish if not conserved.

copying such an individual, it is checked whether its instance does not already exist in the population. It could have been chosen through mating selection and remained unaltered in the population. The insertion of these dominating individuals thus happens only when they are not members of the next generation, with the aim of avoiding the introduction of identical prototypes in the population.

Concerning the preservation of the species, the new imposition that the niches are kept occupied by a number of individuals proportional to their resources, which is achieved both within the earlier TSC and the new TSC2, by means of weighted mating selection, represents a mechanism that is not integrated within SCGA. Within the complementary MGA [5], however, it is claimed that the selection mechanism has influence upon the number of found peaks and, as a consequence, two types of selection are chosen. One is the global weighted selection and the other one is the local selection within each subpopulation (nation). In the previous TSC [14], both selection types are employed with the aim of keeping the population properly distributed. No important influence was observed as concerns the results and consequently the more direct option, i.e., global weighted selection, is herein adopted.

As regards the annulment of multiple instances for a seed, this is a very important difference of the novel TSC2 in comparison to the corresponding procedure within either the initial TSC or SCGA.

3) *Determining the Species:* Before referring to subpopula-

Ideea si experimentele

- Introducerea contine afirmatii
- Corpul lucrarii aduce dovezi care sa sustina acele afirmatii sub forma de:
  - Analize
  - Masuratori
  - Comparatii
  - Studii de caz

TABLE IV

BEST/AVERAGE RESULTS OBTAINED IN 30 LHS POINTS, EACH REPLICATED 30 TIMES, FOR FUNCTIONS $F1$–$F5$

Method	Peak Ratio		Basin Ratio		Peak Accuracy		Distance Accuracy	
	Best	Avg.	Best	Avg.	Best	Avg.	Best	Avg.
$F1$, 1 global optimum, 9 local ones								
TSC2	**0.99**	**0.84**	1	0.88	**0.13**	1.84	**0.04**	0.79
CDE	0.88	0.79	0.98	**0.93**	0.52	**1.59**	0.11	**0.41**
TSC [14]	0.85	0.64	0.83	0.66	4.52	7.7	1.29	3.26
NCMA-ES	0.8	0.49	0.9	0.59	1.85	8.89	0.88	3.87
SCGA	0.66	0.18	0.99	0.264	8.74	18.59	0.98	11.56
DFS	0.37	0.16	0.37	0.16	14.46	20.93	5.24	11.52
$F2$, 2 global, 4 local optima								
TSC2	1	**0.77**	1	**0.77**	**6.93e−04**	**2.91**	0.02	2.09
NCMA-ES	1	0.59	1	0.61	1.72e−03	3.9	0.02	3.19
CDE	1	0.75	1	0.76	0.02	3.3	0.1	**1.99**
SCGA	0.96	0.32	1	0.35	0.39	6.37	0.44	7.02
DFS	0.67	0.26	0.67	0.26	4.64	7.27	2.73	6.22
TSC [14]	0.63	0.46	0.66	0.44	3.93	6.18	3.44	6.18
$F3$, 2 dimensions, 1 optimum								
NCMA-ES	1	1	1	1	**4.6e−68**	3.92e−06	**6.48e−35**	5.84e−04
CDE	1	1	1	1	9.47e−40	4.48e−04	1.96e−20	5.25e−03
TSC2	1	1	1	1	5.85e−12	1.81e−07	1.61e−06	9.32e−05
SCGA	1	1	1	1	1.53e−11	2.86e−07	2.41e−06	1.65e−04
TSC [14]	1	1	1	1	2.48e−10	**1.75e−07**	4.9e−06	**9.08e−05**
DFS	1	1	1	1	2.55e−09	4.17e−06	4.23e−05	8.12e−04
$F3$, 10 dimensions, 1 optimum								
CDE	1	**0.83**	1	1	**2.66e−25**	0.11	**4.07e−13**	**0.15**
NCMA-ES	1	0.73	1	1	1.28e−17	**0.08**	2.51e−09	0.19
TSC2	1	0.73	1	1	2.36e−06	0.15	0.001	0.23
TSC [14]	1	0.74	1	1	2.79e−06	0.12	0.003	0.51
SCGA	1	0.72	1	1	1.03e−05	1.43	0.003	0.45
DFS	1	0.72	1	1	3.12e−05	0.14	0.005	0.22
$F4$, 2 dimensions, 1 global optimum/many local ones								
NCMA-ES	1	0.86	1	0.88	**0**	0.19	**9.05e−9**	0.14

Structura

- Abstract (4 fraze sau cca 150-200 de cuvinte)
- Introducere (1 pagina)
- Problema (1 pagina)
- Cercetari similare (1-2 pagini)
- Ideea proprie (2 pagini)
- Detalii rezultate (5 pagini)
- **Concluzii si idei de continuare a cercetarii (0.5 pagini)**

Concluzii si *Future Work*

- Se plaseaza studiul si importanta sa intr-un context mai larg
- Pot include si un mic sumar al lucrarii
  - Focus pe scopul cercetarii si pe rezultatele cele mai importante
- Indicati daca este cazul o idee de continuare a cercetarii care poate fi continuata de autori sau de cititori
- Nu introduceti lucruri noi la concluzii
  - Toate trebuie sa apara in corpul lucrarii

Laborator

- http://ro.wikibooks.org/wiki/LaTeX
  - Formatare.
- Despartirea in silabe:
  - Cum specificati modul de despartire a cuvantului *policlinica* inca din startul documentului?
- Scriere cu italic, bold etc.
- Modificarea dimensiunii textului
- Exponent si indice
- Spatierea liniilor

Laborator

- Liste: itemize, enumerate, description
- Liste imbricate
- Note de subsol si de margine
- URL-uri cu link in pdf
- Comentarii pe mai multe linii