Recent Advances in Fitness Landscape Analysis

Gabriela Ochoa[1] & Katherine Malan[2]
[1]University of Stirling, Stirling, UK
[2]University of South Africa, Pretoria, South Africa
gabriela.ochoa@stir.ac.uk, malankm@unisa.ac.za

http://gecco-2019.sigevo.org/

GECCO '19 Companion, July 13–17, 2019, Prague, Czech Republic

ACM ISBN 978-1-4503-6748-6/19/07.
https://doi.org/10.1145/3319619.3323383

1

Instructors

- Gabriela Ochoa is a Professor in Computing Science at the University of Stirling, Scotland. She holds a PhD from the University of Sussex, UK. Her research interests include evolutionary and heuristic search methods, with emphasis on autonomous search, hyper-heuristics, fitness landscape analysis and visualisation. She is associate editor for Evolutionary Computation and was for IEEE Transactions on Evolutionary Computation. She served as the EiC for GECCO 2017, and is a member of the SIGEVO board.

- Katherine M. Malan is a senior lecturer in the Department of Decision Sciences at the University of South Africa. She holds a PhD from the University of Pretoria, South Africa. Her research interests include fitness landscape analysis and the application of computational intelligence techniques to real-world problems. She regularly reviews for a number of journals in fields related to evolutionary computation, swarm intelligence and operations research.

2

Outline

- Fundamental Concepts of Fitness Landscapes
  - Motivation for analysing fitness landscapes
  - Basics of fitness landscapes
  - Features of fitness & violation landscapes
- Local Optima Networks (LONs)
  - Definition of Nodes & Edges
  - Detecting Funnels
  - Visualisation & Metrics
- Case Studies
  - LONs applied to feature selection for classification
  - Landscape-aware algorithm selection for constraint-handling
- Closing

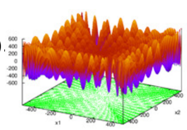
Demo
- Sampling
- Constructing LONs
- Visualisation
- Metrics

Download Materials
lonmaps.com

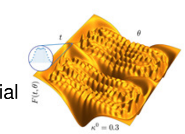
3

Complex Optimisation

- When classical techniques are not feasible & metaheuristics are needed:
  - Problem complexity is too large (not of the required structure for classical techniques, too many variables).
  - When there is no objective function in mathematical form.
  - Objective function exhibits noise or uncertainty.



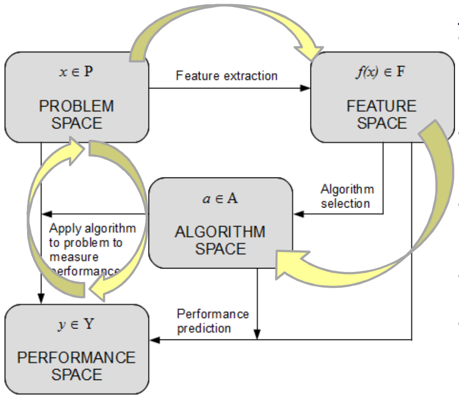
- "Massive optimisation"
  - Large scale optimisation (many dimensions)
  - Any-objective optimisation (single-, multi- many-objective)
  - Cross-domain optimisation (continuous / combinatorial / mixed)
  - Expensive optimisation (costly / simulation-based black-box evaluations)



- Many many metaheuristic approaches…

4

General Algorithm Selection Problem (Rice, 1976)



Trial-and-error approach to finding the best algorithm

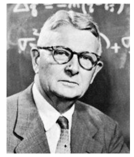
Research areas:

- Feature extraction (characterising problems)
- Analysis and design of problems
- Understanding algorithm behaviour
- Performance prediction
- Algorithm selection

5

Wright's fitness landscape

- ❖ Surface of selective values (Wright, 1932).
- ❖ No axes, units or labels.
- ❖ Commentary 56 years later:
  - *"useless for mathematical purposes"*
  - Aim: provide an intuitive picture of evolutionary processes taking place in higher dimensional space.



1889 - 1988

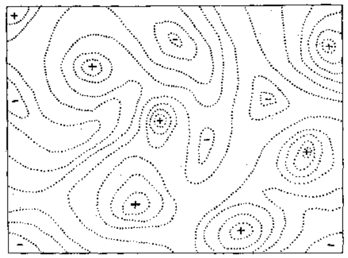
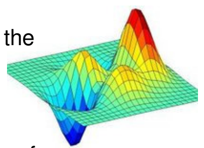


FIGURE 2—Diagrammatic representation of the field of gene combinations in two dimensions instead of many thousands. Dotted lines represent contours with respect to adaptiveness.

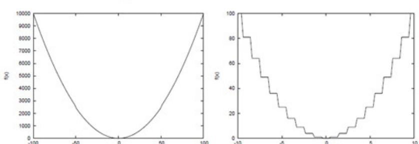
6

Fitness landscapes today

- ❖ We now have (useful) formalised mathematical models.
- ❖ Essential elements: search space, fitness function, notion of neighbourhood or accessibility.
- ❖ Intuitively, a fitness landscape is a visualisation of the terrain capturing how fitness changes between neighbouring solutions.
- ❖ Idea of "valleys", "peaks", "ridges", "plateaus", etc.
- ❖ One fitness function, many fitness landscapes (even for real-valued spaces).
- ❖ Example: Step benchmark function (same function, two different landscapes).



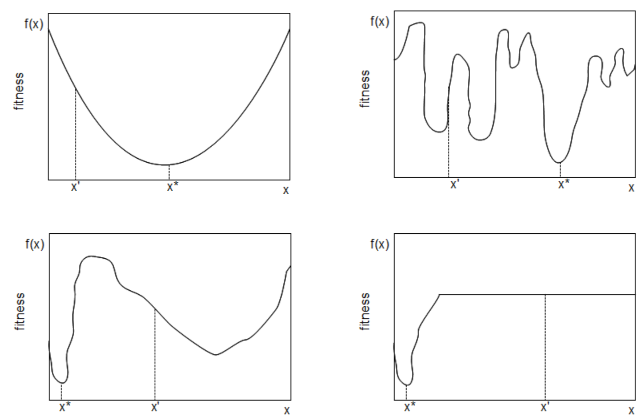
$$f(\mathbf{x}) = \sum_{i=1}^{D} (\lfloor x_i + 0.5 \rfloor)^2$$



(a) Larger domain (b) Smaller domain

7

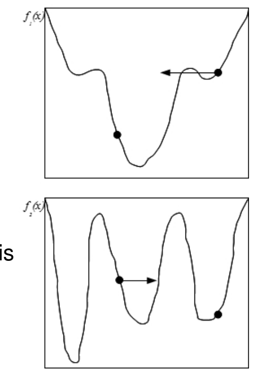
Features of fitness landscapes (1)



8

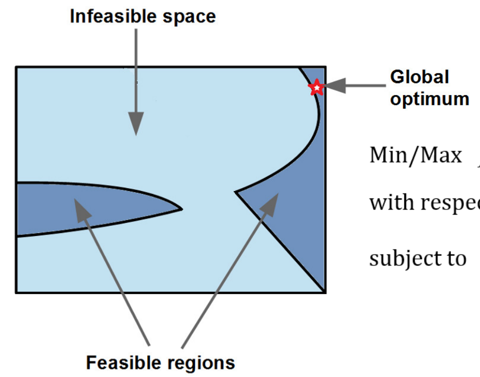
Features of fitness landscapes (2)

- Modality (number of optima) is frequently referred to as affecting difficulty, but too simplistic.
- Example landscapes both with three optima.
- Top landscape: global basin is wider and deeper than local basins.
- Bottom landscape: global basin narrow and local basins deep.
- Consider simple PSO with 2 particles: top landscape not deceptive, bottom landscape is deceptive.
- Distribution & relative sizes of basins of attraction more important than modality.
- Funnels vs ruggedness.



9

Constrained optimisation spaces



Min/Max $f(x_1, x_2, \dots, x_n)$

with respect to $x_1, x_2, \dots, x_n$

subject to $g_k(x_1, x_2, \dots, x_n) \begin{Bmatrix} \le \\ = \end{Bmatrix} b_k$

$k = 1, \dots, m$

- How do constraints impact on the landscape?
- Can view the landscape i.t.o fitness or level of violation.

10

Fitness and Violation Landscapes

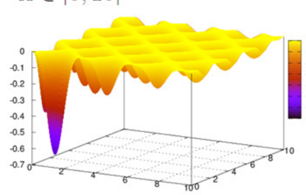
- CEC 2010 Benchmark suite, problem C01:

$$\text{Min } f(\mathbf{x}) = -\left|\frac{\sum_{i=1}^{D}\cos^4(z_i) - 2\prod_{i=1}^{D}\cos^2(z_i)}{\sqrt{\sum_{i=1}^{D} i z_i^2}}\right|, \ \mathbf{z} = \mathbf{x} - \mathbf{o}$$

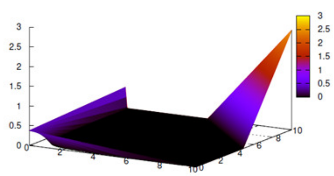
$$g_1(\mathbf{x}) = 0.75 - \prod_{i=1}^{D} z_i \le 0$$

$$g_2(\mathbf{x}) = \sum_{i=1}^{D} z_i - 7.5D \le 0$$

$$\mathbf{x} \in [0, 10]^D$$



Fitness landscape

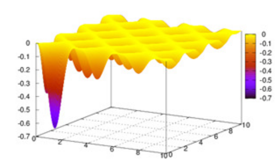


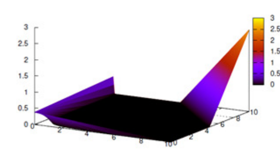
Violation landscape

11

Approaches to constraint handling

- Metaheuristics do not naturally handle constraints, so a constraint-handling technique has to be added on.
- Many different approaches to handling constraints:
  - Avoid infeasible regions (repair / prevention).
  - Use penalties: adapt fitness function to guide search away from infeasible regions.
  - Feasibility ranking: rules of preference using objectives and constraints.
  - Multi-objective optimisation (constraints treated as objective to be minimised).
- What landscape does each approach explore?





12

Which constraint handling technique is better?

- Six approaches: no constraint handling, death penalty, weighted penalty, feasibility ranking, epsilon feasibility ranking, bi-objective.
- Above constraint handling techniques used with evolutionary algorithms performed <u>both the worst and the best</u> on a large set of problems: 142 problem instances (70 combinatorial, 72 continuous).

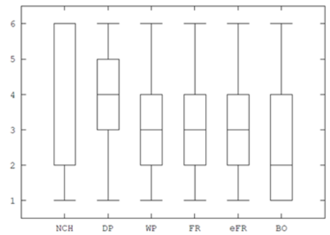
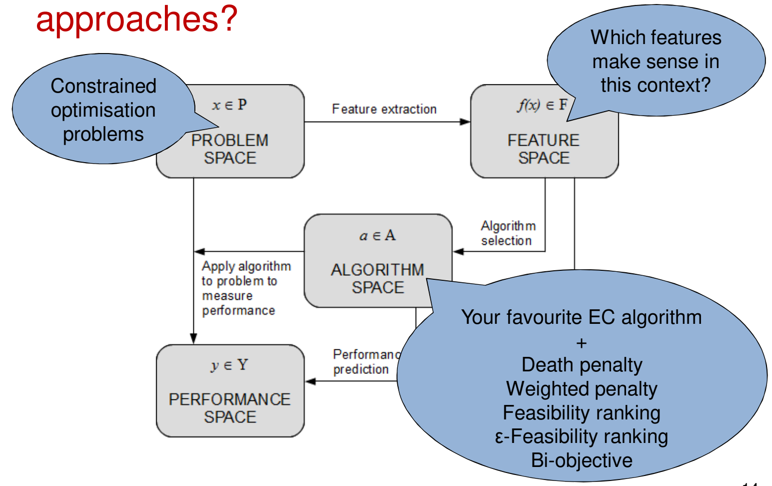


Figure 2: Distribution of algorithm ranks over all problem instances

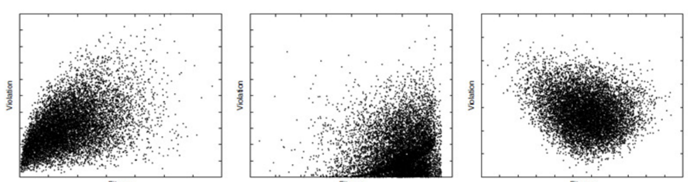
- K.M. Malan and I. Moser. Constraint handling guided by landscape analysis in combinatorial and continuous search spaces. *Evolutionary Computation*, 27(2), 2019.

Algorithm selection for constraint handling approaches?



Features of constrained landscapes

- What proportion of the space is feasible?
- How disjoint are the feasible areas?
- How correlated are the fitness and violation landscapes? Do they “pull” in the same direction?
- What proportion of the solutions are both feasible and fit?



(d) C03 problem in 10D (e) C04 problem in 10D (f) C13 problem in 10D

Sampling to characterise features

Image: pixabay

- Some features require just a random sample, without neighbourhood (e.g. ratio of feasible solutions).
- To characterise how disjoint the feasible areas are, we use a “walk” through the search space.
- Can be random or directed in some way, depending on the purpose of the study.
- Multiple hill climbing walks:
  - More representative of the search process than random walks.
  - Ensures that the sample contains good quality solutions (especially important when some problems have an over-representation of poor solutions).
- Initial solutions must be distributed across the space:
  - Use Latin-hypercube sampling in continuous spaces to prevent over-sampling of the centre of the search space.

Metrics for constrained landscapes

- Feasibility ratio (FsR) estimates the size of the feasible space in relation to the overall search space.
- Given a sample of n solutions, if there are n_f feasible solutions, then

$$\mathrm{FsR} = \frac{n_f}{n}$$

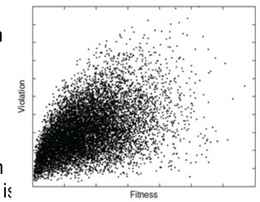
- Ratio of feasible boundary crossings (RFBx) quantifies how disjoint the feasible areas are.
- Given a sequence of n solutions, $x_1, x_2, \ldots, x_n$ obtained by a walk through the search space, a binary string $\boldsymbol{b} = b_1, b_2, \ldots, b_n$ is defined such that $b_i = 0$ if x_i is feasible and $b_i = 1$ if x_i is infeasible, then

$$\mathrm{RFB\times} = \frac{\sum_{i=1}^{n-1} cross(i)}{n-1} \quad \text{where} \quad cross(i) = \begin{cases} 0 \text{ if } b_i = b_{i+1} \\ 1 \text{ otherwise.} \end{cases}$$

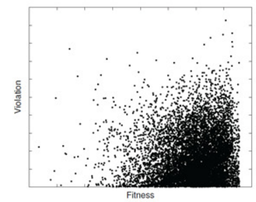
K.M. Malan, J.F. Oberholzer, and A.P. Engelbrecht, A.P. Characterising Constrained Continuous Optimisation Problems. IEEE CEC 2015.

Metrics for constrained landscapes (2)

- Fitness violation correlation (FVC):
  - Based on a sample of solutions resulting in fitness-violation pairs, the FVC is the Spearman's rank correlation coefficient between the fitness and violation values.



- Ideal zone (IZ) metrics:
  - Based on the scatterplot of fitness-violation pairs of a sample of solutions: "ideal zone" is the bottom left corner for a minimisation problem.
  - A small proportion of solutions in the ideal zone could indicate narrow basins of attraction in a penalised landscape.
  - 25_IZ: proportion of points below the 50% percentile for both fitness and violation.
  - 4_IZ: proportion of points below the 20% percentile for both fitness and violation.



Can features predict success / failure?

- 142 problem instances (half combinatorial, half continuous):
  - Samples generated using multiple hill climbing walks from random starting positions.
  - Size: 1% of computational budget used for solving.
  - Samples used as the basis for five landscape metrics.
  - Spearman's correlation coefficient between metrics and success/failure.

	FsR	RFBx	FVC	4_IZ	25_IZ
NCH	0.13	0.26	0.40	0.26	0.23
DP	0.41	0.40	-0.12	-0.33	-0.33
WP	-0.19	-0.10	0.31	0.21	0.32
FR	-0.09	-0.09	0.08	0.01	0.02
ϵFR	-0.04	-0.06	0.10	0.05	0.03
BO	0.17	0.10	-0.10	0.13	-0.15

- DP failure associated with low feasibility
- NCH does well when fitness and violation are correlated.
- Others have weak correlations.

- K.M. Malan and I. Moser. Constraint handling guided by landscape analysis in combinatorial and continuous search spaces. *Evolutionary Computation*, 27(2), 2019.

Dig deeper – split dataset

Table 4: Spearman's correlation coefficients between algorithm performances (success/failure) and landscape metrics for the split datasets.

(a) Correlations for dataset DS_NoFeas.

	FVC	4_IZ	25_IZ
NCH	0.43	0.49	0.40
DP	-0.31	-0.23	-0.31
WP	0.02	0.00	0.02
FR	-0.06	-0.10	-0.06
ϵFR	-0.01	-0.04	0.01
BO	0.35	0.57	0.30

(b) Correlations for dataset DS_Feas.

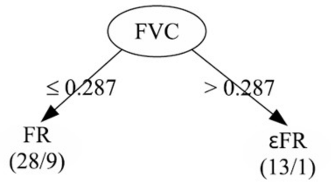
	FsR	RFBx	FVC	4_IZ	25_IZ
NCH	-0.16	0.14	0.51	0.36	0.37
DP	0.31	0.30	0.01	-0.28	-0.22
WP	-0.19	-0.07	0.43	0.24	0.36
FR	0.14	0.13	0.09	-0.06	-0.08
ϵFR	0.08	0.02	0.11	0.03	-0.05
BO	-0.11	-0.30	-0.28	0.15	-0.14

When there is no measurable feasibility (FsR = 0):
- BO correlates positively with 4_IZ metric

When there is measurable feasibility (FsR > 0):
- WP correlates positively with FVC
- BO correlates negatively with RFBx

Decision tree induction sometimes useful

- Predicting when one technique more suited than another technique.



Testing accuracy:
71% using 10-fold cross validation

Prediction in fuzzy terms: *When FsR is zero, if FVC is high, then εFR will probably perform better than FR.*

If there are very few feasible solutions, it is not helpful to ignore the constraint violation early on in the search process (unless the fitness and violation are correlated)

21

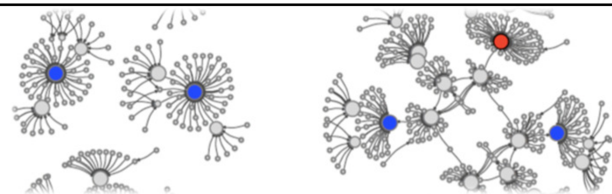
The next steps

- Your initial investigation shows some links between problem features and algorithm performance.
- Next step: design a landscape-aware approach that exploits this knowledge ….

Image: pixabay

- See the last case study of the tutorial.

22

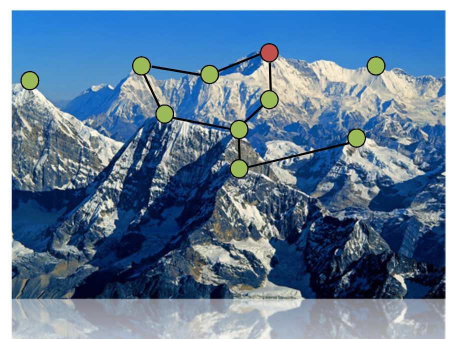


- Overview
- Definition of nodes
- Definition of Edges: basin, escape, monotonic, crossover
- Visualisation & Metrics

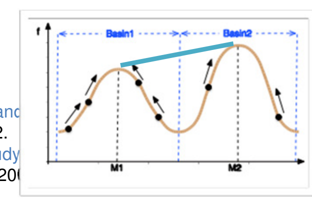
LOCAL OPTIMA NETWORKS

23

Local Optima Networks (LONs)



- Nodes: local optima according to a hill-climbing heuristic
- Edges: possible transitions between optima



- P. K. Doye. The network topology of a potential energy land a static scale-free network. *Physical Review Letter*, 2002.
- G. Ochoa, M. Tomassini, S. Verel, and C. Darabos. A study landscapes' basins and local optima networks. *GECCO* 20

LON original model

❖ Space S, Neigborhood $N(s)$, fitness $f(s)$

❖ $h(s)$ stochastic operator that associates each solution s to its local optimum (Alg. 1)

❖ The basin of attraction of a local optimum $l_i \in L$ is the set $B_i = \{s \in S \mid h(s) = l_i\}$

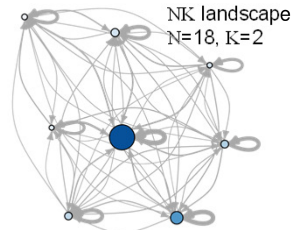
❖ Nodes (L). A local optima is a solution l such that $\forall\ s \in N(s), f(s) \leq f(l)$

❖ Basin Edges (E). Two local optima are connected if their basins of attraction intersect. At least one solution within a basin has a neighbour within the other basin.

❖ LON Model. Directed graph LON = (L, E)

Algorithm 1: Best-improvement local search

```
Choose initial solution s ∈ S
repeat
  choose s' ∈ N(s), f(s') = max_{x∈N(s)} f(x)
  if f(s) ≤ f(s') then
    s ← s'
  end if
until s is a Local optimum
```



w_{ij} proportion of transitions from solutions $s \in Bi$ to solutions $s' \in Bj$

Escape edges

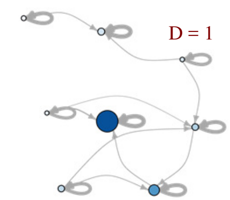
❖ Account for the chances of escaping a local optimum after a controlled mutation (e.g. 1 or 2 bit-flips in binary space) followed by hill-climbing

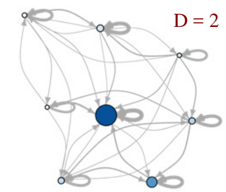
❖ Given a distance function d and integer value D, there is and edge e_{ij} between l_i and l_j if a solution s exists such that $d(s,l_i) \leq D$ and $h(s) = l_j$

❖ w_{ij} cardinality of $\{s \in S \mid d(s,l_i) \leq D$ and $h(s) = l_j\}$

❖ Sampled networks. There is an edge e_{ij} between l_i and l_j if l_j can be obtained after applying a *perturbation* to l_i followed by hill-climbing. Weights are estimated by the sampling process.

NK landscape
N=18, K=2



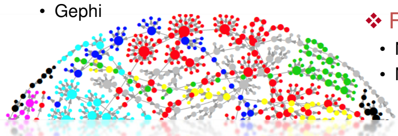


Complex network tools

Visualisation

❖ Force directed layout
- Position nodes in 2D
- Edges of similar length
- Minimise crossings
- Exhibit symmetries
- Example algorithms
  - Fruchterman & Reingold
  - Kamada & Kawai

❖ Software packages
- R igraph
- Gephi

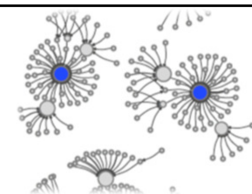


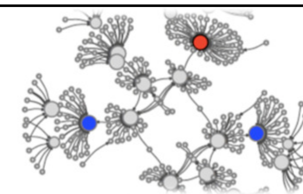
Metrics

❖ Network metrics
- Number of nodes and edges
- Number of global optima
- Weight of self-loops
- Avg. fitness of local optima
- Number of connected components
- Avg. path length to a global optimum
- Centrality (PageRank) of global optima
- Clustering coefficient

❖ Fractal metrics

❖ Funnel metrics
- Number of funnels
- Normalised size of global funnel



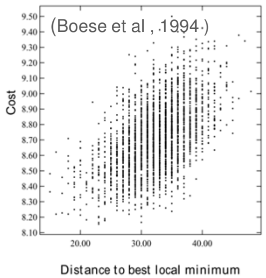


- The Big-Valley Hypothesis
- What is a Funnel?
- Characterization of Funnels with LONs
- Number Partitioning and the Phase transition

FUNNELS

The *big-valley* structure in combinatorial optimisation

- Several studies in the 90s. TSP (Boese et al, 1994), NK landscapes (Kauffman, 1993), graph bipartitioning (Merz & Freisleben, 1998) flowshop scheduling (Reeves, 1999)
- Distribution of local optima is not uniform. Clustered in a big-valley (*globally convex)* structure
- Many local optima, but easy to escape. Gradient at the coarse level leads to the global optimum.



100 cities TSP instance, 2,500 2-opt local optima

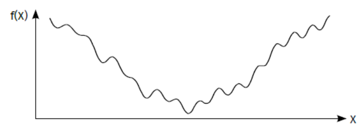


Fig. 1: Depiction of the 'big-valley' structure.

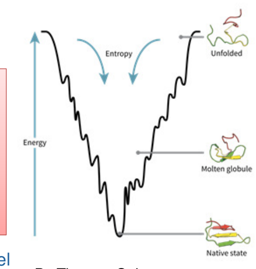
TSP: big-valley. Local optima confined to a small region

29

What is a Funnel?

"A key concept that has arisen within the protein folding community is that of a *funnel* consisting of a set of downhill pathways that converge on a single low-energy minimum."

Doye, J. P. K., Miller, M. A., & Wales, D. J . The double-funnel energy landscape of the 38-atom Lennard-Jones cluster. *Journal of Chemical Physics*, 1999



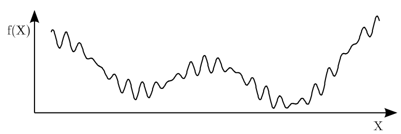
By Thomas Splettstoesser (link)(www.scistyle.com) - Own work

Funnels in continuous optimisation
- Multilevel global structure (Locatelli, 2005)
- *Dispersion* metric (Lunacek &Whitley, 2006, 2008)
- Feature-based detection of (single) funnel structure (Kerschke et al., 2015)

Funnels in combinatorial optimisation
- Related to the big-valley (central-massif) hypothesis (previous slide)
- The big-valley re-visited (Hains, Whitley & Howe, 2011)
- Characterisation of funnels with Local Optima Networks (our contribution)

30

Characterisation of funnels



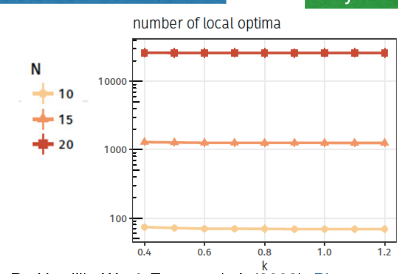
- Funnels can be loosely defined as groups of local optima, which are close in configuration space within a group, but well-separated between groups.
- A funnel conforms a coarse-grained gradient towards a low cost optimum.
- How to characterise funnels more rigorously using LONs?
  - Connected components. Funnels are sub-graphs, connected components within LONs. (EvoCOP, 2016)
  - Communities. Funnels are *communities* within LONs. (GECCO, 2016, 2017)
  - Monotonic sequences. Concept from energy landscapes. Conceptually sound characterisation, incorporating both grouping and coarse-grained gradient. (EvoCOP 2017, 2018; JoH 2017)

31

NPP fitness landscape

What features of the fitness landscape are responsible for the widely different behaviours?

Most fitness landscape metrics are insensitive/oblivious to the easy/hard phase transition!



- Stadler, P., Hordijk, W., & Fontanari, J. (2003). Phase transition and landscape statistics of the number partitioning problem. *Physical Review E*
- K. Alyahya, J. Rowe (2014). Phase Transition and Landscape Properties of the Number Partitioning Problem. *EvoCOP*.

3
2

32

Characterisation of funnels with LONs

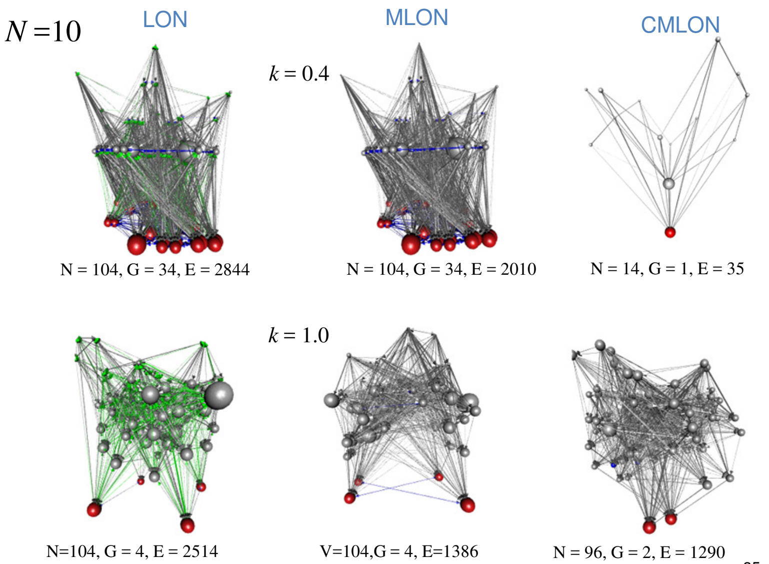
- Monotonic edges. Keep only non-deteriorating edges
  $l_s \rightarrow l_e$, if $f(l_e) \leq f(l_s)$
- Monotonic sequence. Path of connected local optima
  $l_1 \rightarrow l_2 \rightarrow l_3 \ldots \rightarrow l_s$ $f(l_i) \leq f(l_{i-1})$
- Sink. Natural end of the sequence, when there is no adjacent improving local optima
- Definition of Funnel
  - Aggregation of all monotonic sequences ending at the same point (sink).
  - Basin of attraction level of local optima

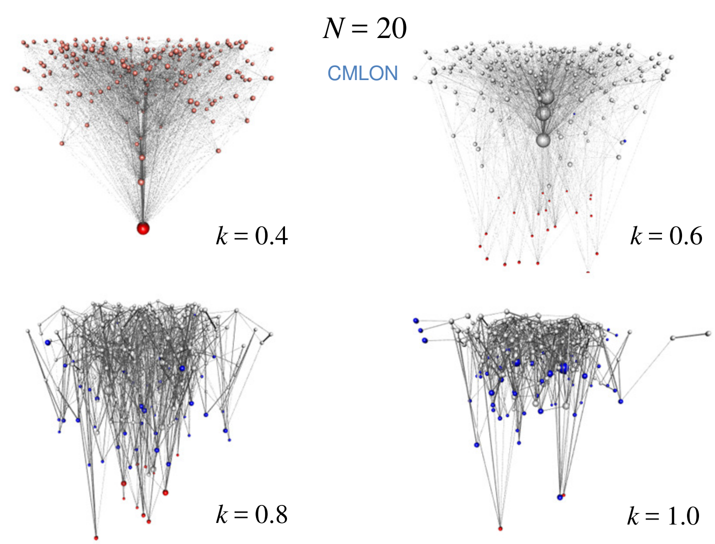
Sink. Node without outgoing edges

$i \leftarrow 0$ $\qquad$ S set of sinks
for $s \in S$ **do**
$\quad fbasin[i] \leftarrow$ breadthFirstSearch(LON, s)
$\quad fbsize[i] \leftarrow$ length($fbasin[i]$)
$\quad i \leftarrow i+1$
end

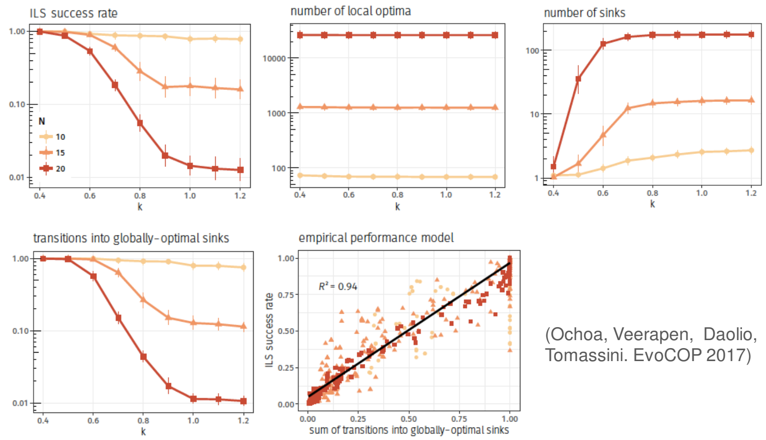
Methodology

- Full enumeration and extraction of LONs
- $N = \{10, 15, 20\}$, k in $[0.4, 1.2]$ step 0.1
- 30 instances for each N and k
- LON. 1-flip local search, 2-flip perturbation $(D = 2)$
- MLON. Monotonic LON, worsening edges pruned
- CMLON. compressed MLON, LON plateaus contracted in a single node
- Empirical search performance: ILS success rate

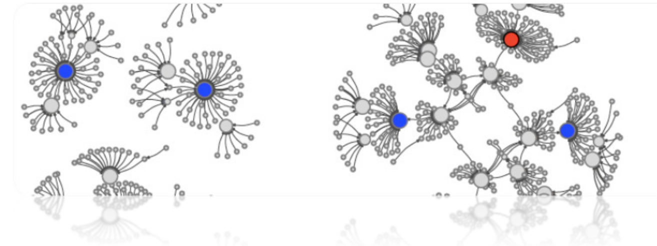




LON metrics & Search Performance



(Ochoa, Veerapen, Daolio, Tomassini. EvoCOP 2017)



- Travelling Salesperson Problem (TSP) and multiple funnels
- Exploiting knowledge of the global structure
  - Increasing perturbation strength
  - Models including recombination
- Feature Selection for Classification

CASE STUDIES

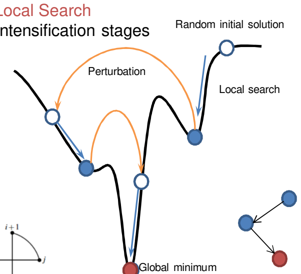
Travelling Salesperson Problem

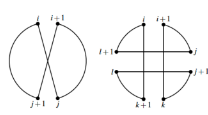
Sampling and constructing LONs with escape edges

Data: I, TSP instance
Result: L, set of local optima,
E, set of escape edges
$L \leftarrow \{\}; E \leftarrow \{\}$
for $i \leftarrow 1$ **to** 1000 **do**
$s_{start} \leftarrow$ initialSolution()
$s_{start} \leftarrow$ LK(s_{start})
$L \leftarrow L \cup \{s_{start}\}$
while $j < 10000$ **do**
$s_{end} \leftarrow$ applyKick(s_{start})
$s_{end} \leftarrow$ LK(s_{end})
$j \leftarrow j + 1$
if $Objective(s_{end}) \leq Objective(s_{start})$ **then**
$L \leftarrow L \cup \{s_{end}\}$
$E \leftarrow E \cup \{(s_{start}, s_{end})\}$
$s_{start} \leftarrow s_{end}$
$j \leftarrow 0$
end
end
end

TSP heuristic, Chained Lin-Kernighan
(Martin, Otto, Felten, 1992)

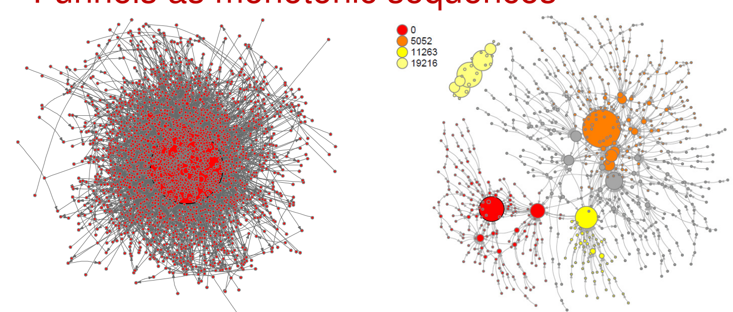
- Form of Iterated Local Search
- Diversification & Intensification stages





- Nodes. Lin-Kernighan
- Edges. Double-bridge

TSP Synthetic Instances
Funnels as monotonic sequences



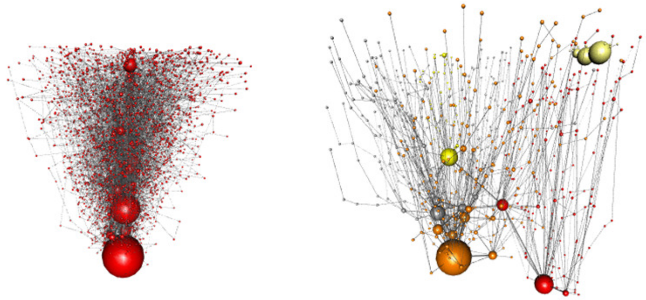
C755 Clustered Cities
Funnels: 1, Success: 100%

E755 Uniform Cities
Funnels: 4, Success: 13%

DIMACS random instances

(Ochoa & Veerapen, JoH 2018)

TSP Synthetic Instances



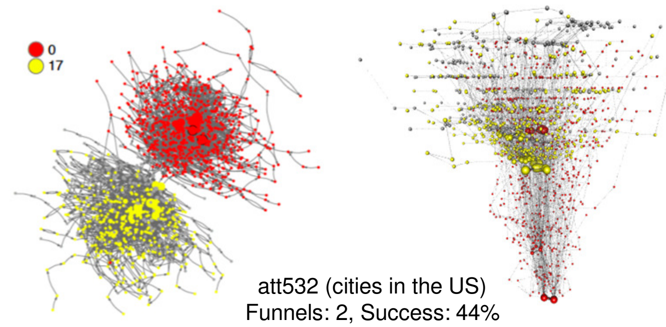
C755 Clustered Cities
Funnels: 1, Success: 100%

E755 Uniform Cities
Funnels: 4, Success: 13%

DIMACS random instances

Same layout, 3D projection where z coordinate is fitness

TSPLIB City Instance att532



att532 (cities in the US)
Funnels: 2, Success: 44%

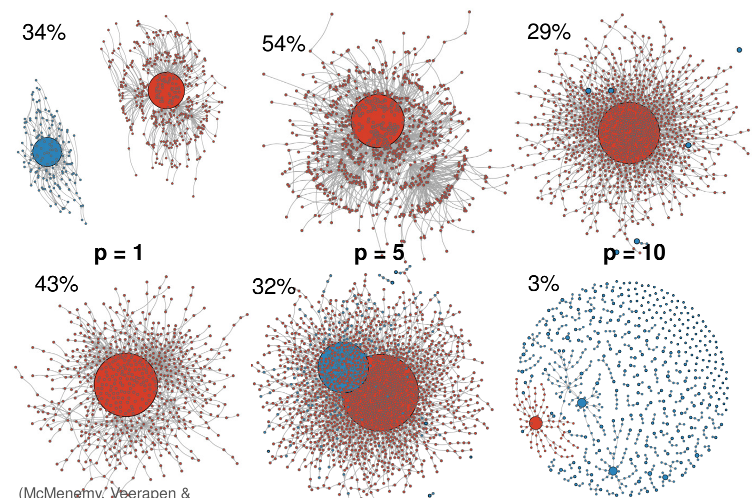
2D layout and 3D projection where z coordinate is fitness

Exploiting knowledge of the global structure

- Instances of several combinatorial optimisation problems have a multi-funnel structure
- Sub-optimal funnels act as traps to the search process
- Can we devise mechanisms for escaping sub-optimal funnels?
  - Restarts
  - Stronger perturbation in ILS implementations
  - Crossover

Increasing perturbation strength

TSP Uniform E755



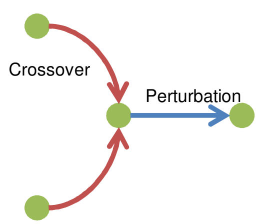
(McMenemy, Veerapen & Ochoa. EvoCOP 2018)

TSP Clustered C755

LONs for Hybrid EAs

- Hydrid EAs which incorporate a local search component to generate local optima.
- Two types of edges
  - Crossover (followed by local search)
  - Perturbation (followed by loca

Grey-box Optimisation
DRILS Deterministic Recombination (PX) + ILS



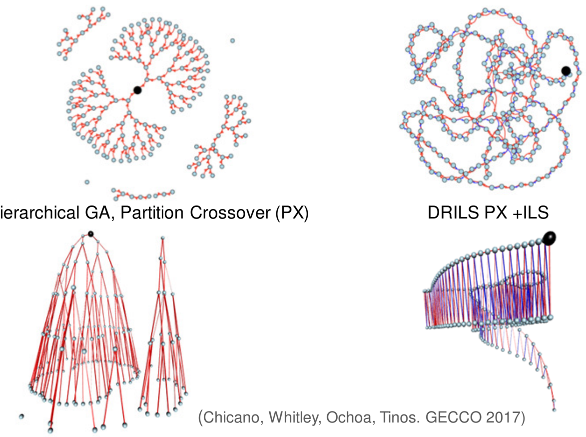
Algorithm 1 DRILS

```
current ← hillClimber(random());
while not stopping condition do
    next ← hillClimber (perturb(current));
    child ← PX(current, next);
    if child = current or child = next then
        current ← next;
    else
        current ← hillClimber(child);
    end if
end while
```

(Chicano, Whitley, Ochoa, Tinos. GECCO 2017)

Contrasting LONs from two solving methods

- Grey-box hybrid EA, 1 million variables NK



(Chicano, Whitley, Ochoa, Tinos. GECCO 2017)

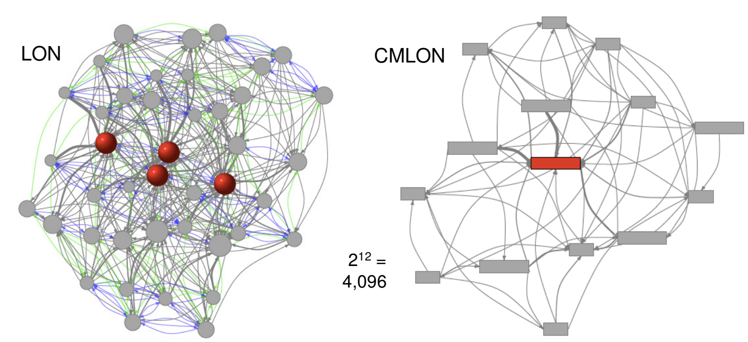
Binary feature selection for classification

- 7 real-world datasets from UCL ML & Open ML, with up to 16 features
- Search space: binary strings, subset of features, 2^n
- Fitness function: classification accuracy
  - *k*-Nearest Neighbour (*k*-NN) algorithm
  - *Cohen's Kappa* statistic
- Fully enumerated LONs (1-bitflip, Escape $D = 2$)
- To test problem difficulty, 3 features selection methods: filter, *sequential forward selection* (SFS) and simple GA wrapper

(Mostert, Malan, Ochoa, Engelbrecht EvoCOP 2019)

Feature selection: vowel recognition dataset

12 features (numeric - extracted from speech) , 11 classes (vowels of British English)



noptima: 42, *nglobals*: 4
SFS_{pr}: 1.00, GA_{pr}:0.98, GA_{sr}: 0.7

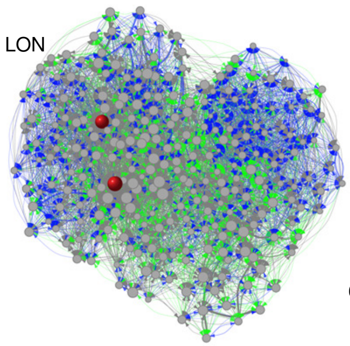
ncoptima: 15, *ncglobal*: 1
neutral: 0.64, *str_global*: 0.50

(Mostert, Malan, Ochoa, Engelbretch EvoCOP 2019)

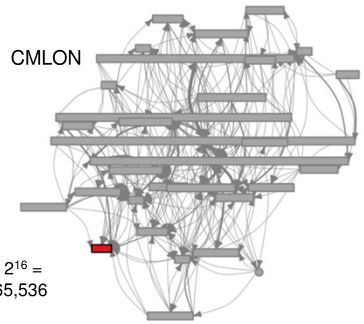
Feature selection: zoo dataset

16 features (B: feathers, aquatic, etc.; I: no.legs), 7 classes (type of animal: mammal, bird, etc.)

LON



CMLON



2^{16} = 65,536

noptima: 4,938, *nglobals*: 2
SFS$_{pr}$: 0.83, GA$_{pr}$:0.99, GA$_{sr}$: 0.13

ncoptima: 1,342, *ncglobal*: 1
neutral: 0.73, *str_global*: 0.001

(Mostert, Malan, Ochoa, Engelbrecht EvoCOP 2019)

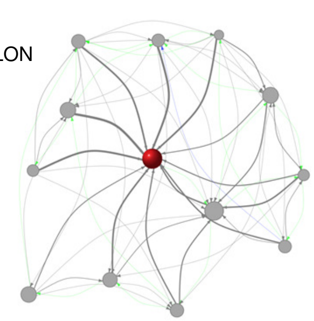
Feature selection: irrelevant features

- Four global optima vowel dataset
  - 0011111100100
  - 0111111100100
  - 1011111100100
  - 1111111100100
- The first two features seem irrelevant
- Checking the dataset confirms that!

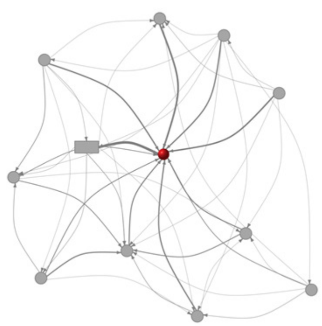
Speaker	Sex	Feature_0	Feature_1	Feature_2	Feature_3	Feature_4	Feature_5	Feature_6	Feature_7	Feature_8	Feature_9	Class
Andrew	Male	-3.639	0.418	-0.67	1.779	-0.168	1.627	-0.388	0.529	-0.874	-0.814	**hid**
Andrew	Male	-3.327	0.496	-0.694	1.365	-0.265	1.933	-0.363	0.51	-0.621	-0.488	**hId**
Andrew	Male	-2.12	0.894	-1.576	0.147	-0.707	1.559	-0.579	0.676	-0.809	-0.049	**hEd**
Andrew	Male	-2.287	1.809	-1.498	1.012	-1.053	1.06	-0.567	0.235	-0.091	-0.795	**hAd**
Andrew	Male	-2.598	1.938	-0.846	1.062	-1.633	0.764	0.394	-0.15	0.277	-0.396	**hYd**
Andrew	Male	-2.852	1.914	-0.755	0.825	-1.588	0.855	0.217	-0.246	0.238	-0.365	**had**

Feature selection: vowel recognition dataset after removing irrelevant features

LON



CMLON



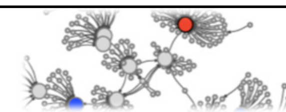
2^{10} = 1,024

noptima: 13, *nglobals*: 1
SFS$_{pr}$: 1.00, GA$_{pr}$:0.98, GA$_{sr}$: 0.8

ncoptima: 12, *ncoptima*: 1
neutral: 0.08, *str_global*: 0.59

(Mostert, Malan, Ochoa, Engelbrecht EvoCOP 2019)

LONs



Conclusions

- More accessible (visual) approach to heuristic understanding
- Global structure impacts search
- Real-world problems are neutral and have multiple-funnels

Contributions

- Sampling & Visualisation
- Characterisation of funnels
- New LON metrics can improve performance prediction.
- Using knowledge to select/configure algorithms

lonmaps.com - Website with LON resources

Image: pixabay

CASE STUDY 2: LANDSCAPE-AWARE CONSTRAINT HANDLING

53

Landscape aware constraint handling

Overall approach:

- ❖ Benchmark suite of training instances: characterise instances using landscape analysis based on sampling.
- ❖ Solve the problem instances using different constraint handling techniques (CHTs) with the same base algorithm and rank the performances of CHTs on training instances.
- ❖ Perform data mining on training set to model the relationship between problem features and winning algorithm approaches.
- ❖ Formulate high-level rules for selecting appropriate CHTs.
- ❖ Take a different set of problem instances as the test set.
- ❖ Solve them by switching to the CHT that is predicted to be the best, given the landscape characteristics experienced during search (online landscape analysis).
- ❖ Compare the landscape-aware approach to the individual CHTs.

- K.M. Malan, Landscape-aware Constraint Handling Applied to Differential Evolution. TPNC 2018, LNCS 11324, pp. 176-187.

54

Benchmark suite

- ❖ Limited to real-valued minimisation problems:

$$\text{Minimise } f(\mathbf{x}), \quad \mathbf{x} = (x_0, x_1, \ldots, x_{n-1}) \in \mathbb{R}^n \qquad (1)$$

$$\text{subject to } \begin{array}{ll} g_i(\mathbf{x}) \leq 0, & i = 1, \ldots, p \\ h_j(\mathbf{x}) = 0, & j = p+1, \ldots, m \end{array} \qquad (2)$$

- ❖ Equality constraints re-expressed as inequality constraints ($\varepsilon = 10^{-4}$):

$$|h_j(\mathbf{x})| - \epsilon \leq 0, \quad j = p+1, \ldots, m \qquad (3)$$

- ❖ CEC 2010 Competition on Constrained Real-Parameter Optimization:
  - 18 problems, scalable to any dimension.
  - Training data set: nine odd numbered functions in 5D, 10D, 15D, 20D, 25D and 30D (54 instances).
  - Testing data set: nine even numbered functions in 5D, 10D, 15D, 20D, 25D and 30D (54 instances).

55

Base algorithm: Differential Evolution

- ❖ Differential evolution:
  - Established, popular population-based metaheuristic for real-valued optimisation.
  - Performed well in CEC Constrained Real-Parameter Optimisation competitions: CEC 2006, CEC2010.

1st	**ε_DE**
2nd	**DMS-PSO**
3rd	**SaDE, MDE**
5th	**PCX**
6th	**MPDE**
7th	**DE**
8th	**jDE-2**
9th	**GDE, PESO+**

Rank	Algorithm
1st	εDEg
2nd	ECHT
3rd	jDEsoco
4th	DCDE
5th	Co-CLPSO
6th	IEMA
7th	DE-VPS
8th	CDEb6e6rl
9th	RGA
10th	E-ABC
11th	MTS
12th	sp-MODE

56

Feature extraction of training set

- 54 training instances characterised based on samples.
- Approach to sampling:
  - Sample size: 200 x D (1% of computational budget for solving the problem) generated for each instance using multiple hill climbing walks.
  - From a random initial position, neighbours sampled (from a Gaussian distribution with mean = current position and std deviation = 5% of the range of domain of search space).
  - Walk terminated if no better neighbour found after sampling 100 random neighbours.
- Sample used as the basis for calculating five landscape metrics: FsR, RFBx, FVC, 25_IZ, 4_IZ.

57

Constraint handling techniques

Base algorithm: *DE/rand/1*, with uniform crossover, a population size of 100, a scale factor of 0.5, and a crossover rate of 0.5.

1. Weighted Penalty:
   - Combine constraint violation as a penalty in the objective function (50% penalty, 50% objective value).
2. Feasibility Ranking (Deb, 2000):
   - Two feasible solutions compared on objective value
   - Feasible solution always preferred to an infeasible solution
   - Two infeasible solutions compared by level of constraint violation.
3. ε-Feasibility Ranking (Takahama & Sakai, 2006):
   - Like Deb's rules, but with a tolerance (ε) to constraint violations that reduces over time.
4. Bi-objective:
   - Constraint violation treated as a 2^{nd} objective.
   - Non-dominated sorting of NSGA II (Deb et al., 2002).

58

Measuring performance on training set

- Classic DE run on training instances with each CHT 30 times.
- Computational budget: 20 000 x D.
- A run of an algorithm regarded as feasible if a feasible solution found within the budget of function evaluations.
- Two algorithms compared using CEC2010 competition rules:
  - If two algorithms have different success rates, the algorithm with the higher success rate wins.
  - If two algorithms have the same success rate > 0, the algorithm with the superior mean fitness value of feasible runs wins.
  - If two algorithms have success rate = 0, the algorithm with the lowest mean violation wins.

Image: pixabay

59

Example performance on training set

CEC 2010 problem C01 in 15 dimensions				
	Success rate	Mean fitness (feasible runs)	Mean violation	Algorithm rank
WP	0.233	-0.7824	0.2125	3
FR	1	-0.7815	0	2
ϵFR	1	-0.7820	0	1
BO	0	n/a	0.3750	4
CEC 2010 problem C09 in 5 dimensions				
	Success rate	Mean fitness (feasible runs)	Mean violation	Algorithm rank
WP	1	0.2561	0	2
FR	0	n/a	0.4117	4
ϵFR	0	n/a	0.2637	3
BO	1	0.0000	0	1

60

Deriving CHT selection rules

- Training dataset for a classification problem:
  - 54 problem instances with 5 features (landscape metrics based on sampling).
  - Binary performance class for each algorithm: the best or not.
- Use data mining to predict when each CHT will be the best:
  - C4.5 algorithm used to induce decision trees using the full training dataset.
  - Rules extracted for predicting when each algorithm is expected to be the best.

1. WP is predicted to be the best when (4_IZ > 0.006) AND ((FsR > 0) OR (FsR $= 0$ AND FVC ≤ 0.06)).
2. FR is predicted to be the best when (25_IZ ≤ 0.259) AND (RFBx ≤ 0.083).
3. ϵFR is predicted to be the best when FsR > 0.28.
4. BO is predicted to be the best when (FVC > 0.28) AND (FsR $= 0$).

61

Online landscape analysis

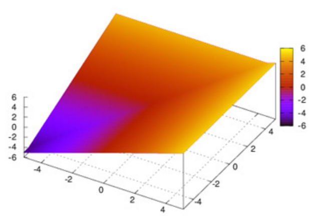
- Landscape information collected during search:
  - Path of each individual in the population is treated as a "walk": position, objective value & constraint values stored in a queue.
  - With each new generation, if the new position differed from previous solution, the new child solution appended to the walk of that individual.
  - *OLA_limit*: parameter for queue length.
- Switching constraint handling using general rules (derived from data mining):
  - After a set number of iterations (*SW_freq* parameter), the landscape metrics are calculated.
  - Based on the landscape profile, the CHT is switched to the predicted best strategy (using the rules derived previously through data mining).

62

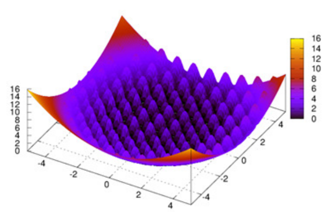
Sample run

Table 3. Data from a sample run of the LA approach on problem C02 in 5D

Iteration	Online Landscape Metrics					Selected Strategy
	FsR	RFBx	FVC	25_IZ	4_IZ	
50	0	0	0.553	0.335	0.017	BO
100	0	0	-0.810	0.063	0	FR
600	0.001	0.001	-0.872	0.050	0	FR



Fitness landscape



Violation landscape

63

Results

- Performance of six constraint handling approaches on 54 test problem instances (rank: 1 – 6):
  - Four base constraint handling techniques: WP, FR, εFR, BO.
  - RS: Random switching between above 4 techniques.
  - LA: Landscape-aware switching based on online landscape features.
- Parameters: *SW_freq* = 10 x D, *OLA_limit* = 10 x D.

Strategy	Mean Rank	Best Performing		Worst Performing	
WP	3.44	**16 instances**	**(30%)**	14 instances	(26%)
FR	3.69	7 instances	(13%)	8 instances	(15%)
ϵFR	3.59	3 instances	(6%)	**0 instances**	**(0%)**
BO	4.54	9 instances	(17%)	32 instances	(59%)
RS	3.19	9 instances	(17%)	**0 instances**	**(0%)**
LA	**2.46**	15 instances	(28%)	**0 instances**	**(0%)**

64

Case study conclusion

- There is value in utilising a range of constraint-handling techniques.
- Proposed switching technique:
  - Pre-processing landscape analysis step to derive rules for predicting when each constraint handling technique will perform the best.
  - Rules applied during search using features extracted from the search path (no additional sampling or fitness evaluations needed).
- Results show that the proposed landscape-aware approach performed better than the constituent approaches when used in isolation.
- Similar approach to landscape-aware search can be used in other contexts.

65

Tutorial conclusion

- Fitness landscape analysis has come a long way in the last 10 years
  - Different perspectives of landscapes: local scale (e.g. ruggedness), global scale (e.g. funnels)
  - Different landscapes (e.g. fitness and violation landscapes).
- We showed how local optima networks can be used
  - To visualise global structure
  - To characterise funnels
- Case studies demonstrated
  - Using LONs to analyse funnels of TSP
  - Exploiting knowledge of global structure to configure search algorithms.
  - Using LONs to gain insight into the feature selection problem
  - How Rice's general algorithm selection framework can be used to implement landscape-aware search in the context of constraint handling techniques for evolutionary algorithms

66

References (1)

- K. Alyahya and J. E. Rowe, Phase Transition and Landscape Properties of the Number Partitioning Problem," *Evolutionary Computation in Combinatorial Optimisation*, EvoCOP, Springer Berlin Heidelberg, 2014, pp. 206–217.
- D. L. Applegate, R. E. Bixby, V. Chvátal, and W. J. Cook, *The Traveling Salesman Problem: A Computational Study*. Princeton University Press, 2006.
- A.-L. Barabási and M. Pósfai, *Network Science*. Cambridge University Press, 2016.
- K. D. Boese, A. B. Kahng, and S. Muddu, A new adaptive multi-start technique for combinatorial global optimizations, *Operations Research Letters*, vol. 16, no. 2, pp. 101–113, Sep. 1994.
- F. Chicano, D. Whitley, G. Ochoa, R. Tinos (2017) Optimizing One Million Variables NK Landscapes by Hybridizing Deterministic Recombination and Local Search. Genetic and Evolutionary Computation Conference, GECCO 2017: 753-760.
- J. P. K. Doye, Network Topology of a Potential Energy Landscape: A Static Scale-Free Network, *Phys. Rev. Lett.*, vol. 88, no. 23, p. 238701, May 2002.
- J. P. K. Doye, M. A. Miller, and D. J. Wales, The double-funnel energy landscape of the 38-atom Lennard-Jones cluster, *The Journal of Chemical Physics*, vol. 110, no. 14, pp. 6896–6906, Apr. 1999.

67

References (2)

- K. Helsgaun, An effective implementation of the Lin–Kernighan traveling salesman heuristic, *European Journal of Operational Research*, vol. 126, no. 1, pp. 106–130, Oct. 2000.
- S. Herrmann, G. Ochoa, F. Rothlauf, PageRank centrality for performance prediction: the impact of the local optima network model. *J. Heuristics* 24(3): 243-264 (2018)
- S. Herrmann, M. Herrmann, G. Ochoa, and F. Rothlauf, Shaping Communities of Local Optima by Perturbation Strength, *Genetic and Evolutionary Computation Conference*, GECCO, 2017, pp. 266–273.
- S. Herrmann, G. Ochoa, and F. Rothlauf, Communities of Local Optima As Funnels in Fitness Landscapes, in *Proceedings of the Genetic and Evolutionary Computation Conference 2016*, New York, NY, USA, 2016, pp. 325–331.
- S. A. Kauffman, *The Origins of Order: Self-organization and Selection in Evolution.* Oxford University Press, 1993.
- P. Kerschke, M. Preuss, S. Wessing, and H. Trautmann, Detecting Funnel Structures by Means of Exploratory Landscape Analysis, *Genetic and Evolutionary Computation Conference* , GECCO, 2015, pp. 265–272.
- E. D. Kolaczyk, G. Csárdi, Statistical Analysis of Network Data with R, Springer, 2014

68

References (3)

- M. Locatelli, On the Multilevel Structure of Global Optimization Problems, *Comput Optim Applic*, vol. 30, no. 1, pp. 5–22, Jan. 2005.
- M. Lunacek and D. Whitley, The Dispersion Metric and the CMA Evolution Strategy, *Genetic and Evolutionary Computation Conference*, GECCO, 2006, pp. 477–484.
- I. Gent and T. Walsh, Phase Transitions and Annealed Theories: Number Partitioning as a Case Study, in *Proceedings of the 12th European Conference on Artificial Intelligence (ECAI-96)*, 1996, pp. 170–174.
- D. R. Hains, L. D. Whitley, and A. E. Howe, Revisiting the big valley search space structure in the TSP, *J Oper Res Soc*, vol. 62, no. 2, pp. 305–312, Feb. 2011.
- M. Lunacek, D. Whitley, and A. Sutton, The Impact of Global Structure on Search, *Parallel Problem Solving from Nature – PPSN X*, Springer Berlin Heidelberg, 2008, pp. 498–507.
- K.M. Malan. Landscape-aware Constraint Handling Applied to Differential Evolution, TPNC 2018, Lecture Notes in Computer Science, vol 11324, Springer, pp 176-187.
- K.M. Malan and I. Moser, Constraint Handling Guided by Landscape Analysis in Combinatorial and Continuous Search Spaces, *Evolutionary Computation,* vol. 27, no. 2, 2019.
- K.M. Malan, J.F. Oberholzer, and A.P. Engelbrecht, A.P. Characterising Constrained Continuous Optimisation Problems. In *Proceedings of the IEEE Congress on Evolutionary Computation*, May 2015, Sendai, Japan, pp 1351-1358.

69

References (4)

- Mallipeddi, R. and Suganthan, P.N, Problem definitions and evaluation criteria for the CEC 2010 competition on constrained real-parameter optimization, Technical report, Nanyang Technological University, Singapore, 2010.
- O. Martin, S. W. Otto, and E. W. Felten, Large-step markov chains for the TSP incorporating local search heuristics, *Operations Research Letters*, vol. 11, no. 4, pp. 219–224, May 1992.
- P. McMenemy, N. Veerapen, and G. Ochoa, How Perturbation Strength Shapes the Global Structure of TSP Fitness Landscapes, *Evolutionary Computation in Combinatorial Optimization*, EvoCOP, 2018, pp. 34–49.
- W. Mostert, K. M. Malan, G. Ochoa, A. P. Engelbrecht, Insights into the Feature Selection Problem Using Local Optima Networks. *Evolutionary Computation in Combinatorial Optimization*, EvoCOP 2019: 147-162
- G. Ochoa, F. Chicano, R. Tinós, and D. Whitley, Tunnelling Crossover Networks, in *on Genetic and Evolutionary Computation Conference* GECCO, 2015, pp. 449–456.
- G. Ochoa, M. Tomassini, S. Vérel, and C. Darabos, A Study of NK Landscapes' Basins and Local Optima Networks, *Conference on Genetic and Evolutionary Computation,* GECCO, 2008, pp. 555–562.
- G. Ochoa and N. Veerapen, Deconstructing the Big Valley Search Space Hypothesis, *Evolutionary Computation in Combinatorial Optimization* EvoCOP, Springer International Publishing, 2016, pp. 58–73.

70

References (5)

- G. Ochoa and N. Veerapen, Mapping the global structure of TSP fitness landscapes, *J Heuristics* J. Heuristics 24(3): 265-294 (2018)
- G. Ochoa, N. Veerapen, F. Daolio, and M. Tomassini, Understanding Phase Transitions with Local Optima Networks: Number Partitioning as a Case Study, *Evolutionary Computation in Combinatorial Optimization,* EvoCOP, 2017, Springer vol. 10197, pp. 233–248.
- John R. Rice (1976), The Algorithm Selection Problem, *Advances in Computers*, Vol 15, Elsevier, pp. 65-118.
- P. F. Stadler, W. Hordijk, and J. F. Fontanari, Phase transition and landscape statistics of the number partitioning problem, *Phys. Rev. E*, vol. 67, no. 5, p. 056701, 2003.
- N. Veerapen, G. Ochoa, Visualising The Global Structure Of Search Landscapes: Genetic Improvement As A Case Study, *Genetic Programming and Evolvable Machines*19(3): 317-349 (2018)
- N. Veerapen, G. Ochoa, R. Tinós, and D. Whitley, Tunnelling Crossover Networks for the Asymmetric TSP," *Parallel Problem Solving from Nature – PPSN XIV*, Springer International Publishing, 2016, pp. 994–1003.
- S. Wright (1932), The Roles of Mutation, Inbreeding, Crossbreeding, and Selection in evolution, In *Proceedings of the Sixth International Congress on Genetics*, pp. 356–366.

71