Are the artificially generated instances uniform in terms of difficulty?

Aritz Pérez, Josu Ceberio, Jose A. Lozano

Rio de Janeiro
July, 2018

basque center for applied mathematics

eman ta zabal zazu
Universidad del País Vasco
Euskal Herriko Unibertsitatea

Outline

Motivation

- Combinatorial optimization problems (**COP**):

$$\text{instances} \equiv \text{parameters}$$

- Artificially generated **benchmarks**
- Usually **parameters** are generated uniformly at random (**u.a.r**)
- How is the **distribution of the difficulty** of the generated instances?
- Goal: **empirically** analyze the distribution of the difficulty
- Difficulty is algorithm dependent: **local search**

Combinatorial optimization problems (COP)

Objective/fitness function

$$
\begin{aligned}
f \quad : \quad & \Omega \to \mathbb{R} \\
& x \mapsto f(x; \Theta)
\end{aligned}
$$

- In this work: linear ordering problem (**LOP**), flowshop scheduling problem (**FSP**) and quadratic assignment problem (**QAP**)
- Discrete search space, Ω: permutations of size n.
- Parameters, Θ: instance.

Outline

Local Search

Neighborhood

Progresses by moving from one solution to a better neighboring.

- First improvement
- Best improvement (**BI**)

Neighborhoods

$$N: \quad \Omega \to 2^{\Omega}$$
$$x \mapsto N(x)$$

2^{Ω} represents the power set of the search space.

- **Swap**: Perform any exchange of two items in consecutive positions (Kendall's distance one).
- **Interchange**: Perform any exchange of the items in any two positions i and j (Cayley's distance one).
- **Insert**: Move any item from a position i to any position j (Ulam's distance one).

Difficulty of an instance

Difficulty

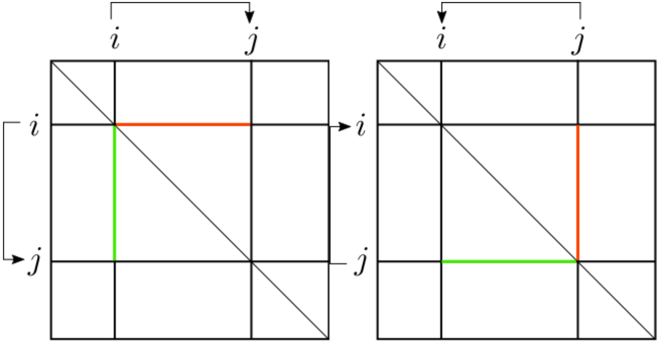
The **probability of not reaching the global optimum** when starting from a solution taken u.a.r.

- Inversely proportional to the size of the **basis of attraction of the global optimum**: number of solutions from which BI reaches the optimum
- Classify instances according to the difficulty: $\mathcal{O}(n!)$ equivalence classes.

Parametric roughness

The **number of parameters that differ** between a solution and its neighbors in the computation of their fitness functions.

- **LOP**: $f(\sigma; B = \{b_{i,j}\}_{n\times n}) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} b_{\sigma_i,\sigma_j}$
- **Insert**



Parametric roughness

	Swap	Interchange	Insert
Size	$\mathcal{O}(n)$	$\mathcal{O}(n^2)$	$\mathcal{O}(n^2)$
LOP	1	$2\lvert i-j\rvert-1$	$\lvert i-j\rvert$
FSP	$m(n-i)$	$m(n-\min(i,j))$	$m(n-\min(i,j))$
QAP	$4(n-1)$	$4(n-1)$	$2n(\lvert i-j\rvert+1)-(\lvert i-j\rvert+1)^2$

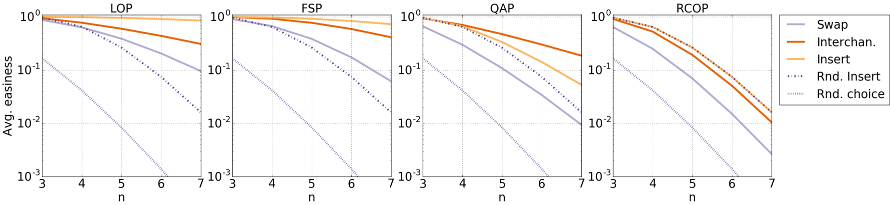
- Related to the **smoothness of the landscapes**.
- Roughness and the **Size** of a neighborhood: distribution of the difficulty.

Outline

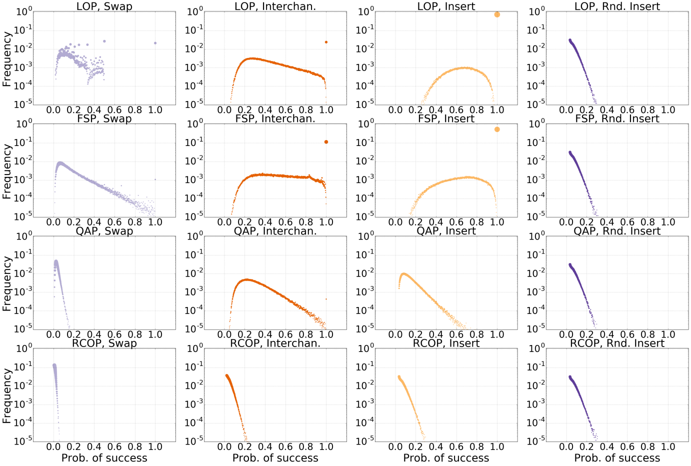
Setup

- COPS: LOP, FSP, QAP.
- Neighborhoods: Swap, Interchange, Insert.
- $5 \cdot 10^5$ instances for each COP.
- **RCOP**: random rankings of solutions.
- **Rand. insert**: relabeling of the insert neighborhood system.

Evolution of the average easiness

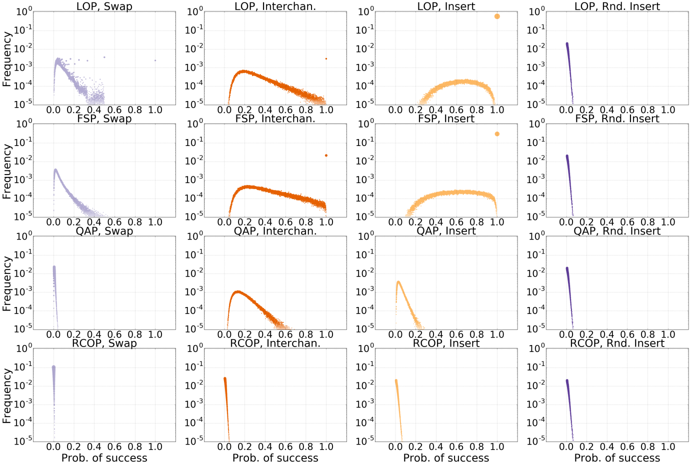


Easiness, $n = 6$



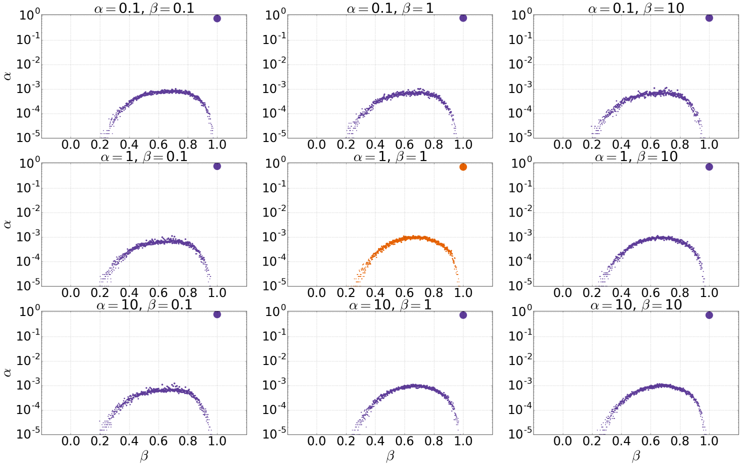
LOP, Swap
LOP, Interchan.
LOP, Insert
LOP, Rnd. Insert
FSP, Swap
FSP, Interchan.
FSP, Insert
FSP, Rnd. Insert
QAP, Swap
QAP, Interchan.
QAP, Insert
QAP, Rnd. Insert
RCOP, Swap
RCOP, Interchan.
RCOP, Insert
RCOP, Rnd. Insert
Frequency
Prob. of success
10^0 10^-1 10^-2 10^-3 10^-4 10^-5
0.0 0.2 0.4 0.6 0.8 1.0

Easiness, $n = 7$


LOP, Swap
LOP, Interchan.
LOP, Insert
LOP, Rnd. Insert
FSP, Swap
FSP, Interchan.
FSP, Insert
FSP, Rnd. Insert
QAP, Swap
QAP, Interchan.
QAP, Insert
QAP, Rnd. Insert
RCOP, Swap
RCOP, Interchan.
RCOP, Insert
RCOP, Rnd. Insert
Frequency
Prob. of success
10^0 10^-1 10^-2 10^-3 10^-4 10^-5
0.0 0.2 0.4 0.6 0.8 1.0

Alternative sampling: *Beta*$(\cdot; \alpha, \beta)$

- LOP + insert, $n = 6$.
- Distribution with many shapes, α and β.



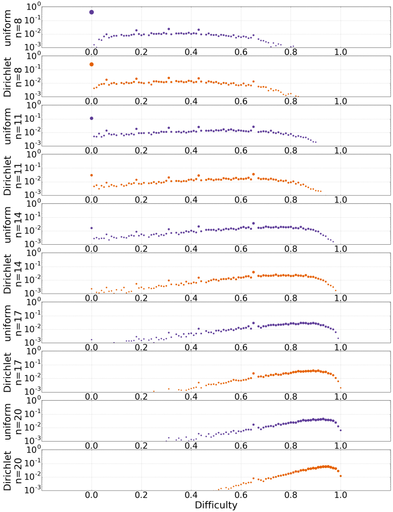
Outline

Conclusions

- Instances are **not uniformly** distributed in terms of difficulty.
- The distribution of the difficulty **depends on the problem and the neighborhood**.
- The distribution of the difficulty seems to be related with the **roughness** and the **size** of a neighborhood.
- A neighborhood with a **low roughness** is desirable

How can we control the difficulty of generated instances?

- LOP + insert: Dirichlet sampling, $n = 8, ..., 20$.



Thanks for your attention!!!

Are the artificially generated instances uniform in terms of difficulty?

Aritz Pérez, Josu Ceberio, Jose A. Lozano

Rio de Janeiro
July, 2018

basque center for applied mathematics

eman ta zabal zazu
Universidad del País Vasco
Euskal Herriko Unibertsitatea