On the Automatic Design of a Representation for Grammar-based Genetic Programming

[best paper at EuroGP 2018]

Eric Medvet and Alberto Bartoli

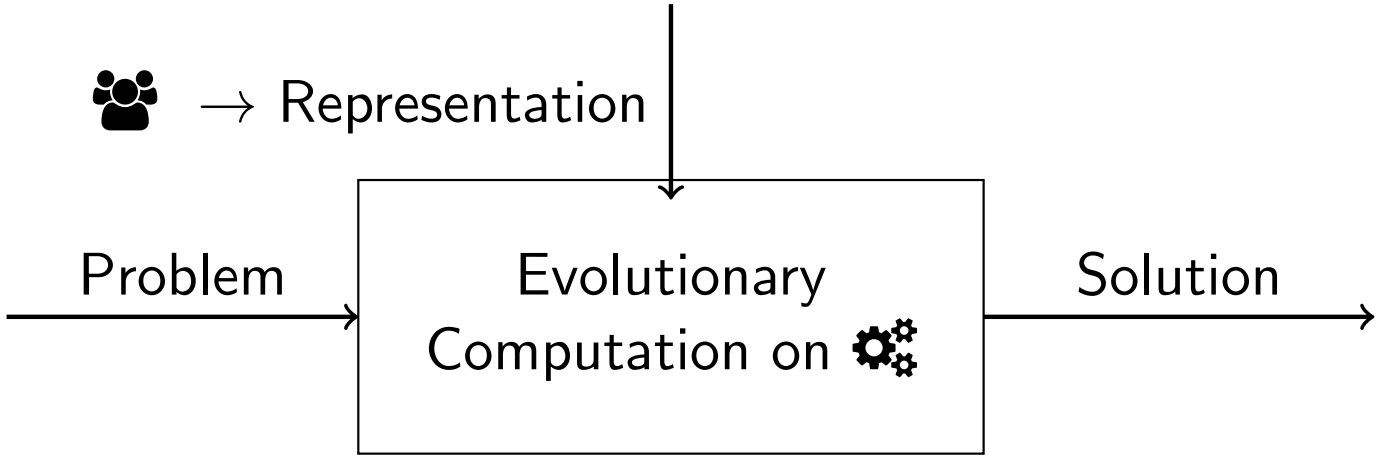
Department of Engineering and Architecture
University of Trieste
Italy

Humies@GECCO, 17/7/2018, Kyoto (Japan)

http://machinelearning.inginf.units.it

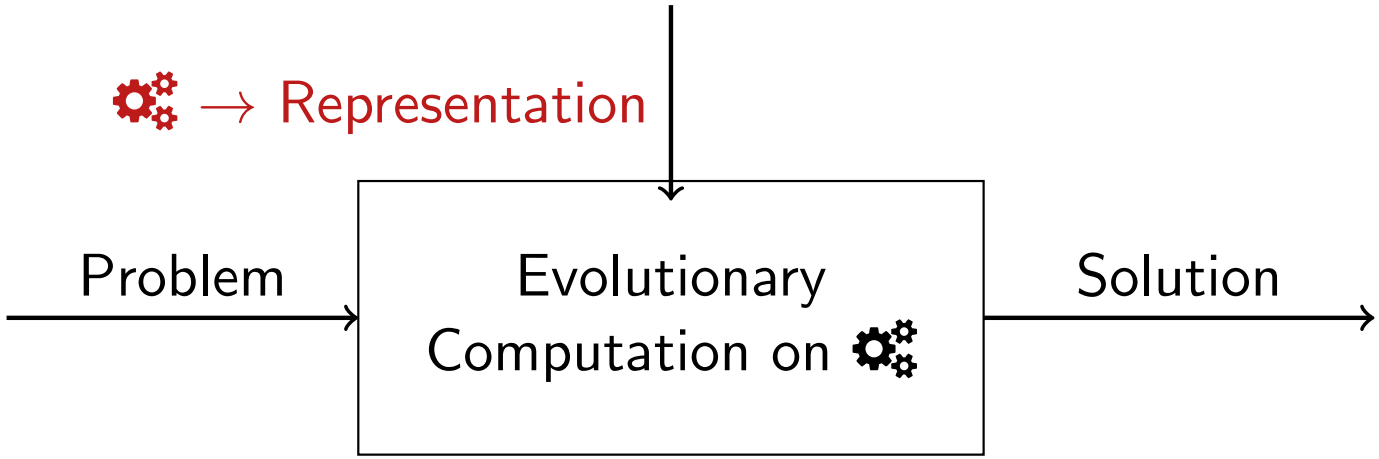
Table of Contents

Individual representation



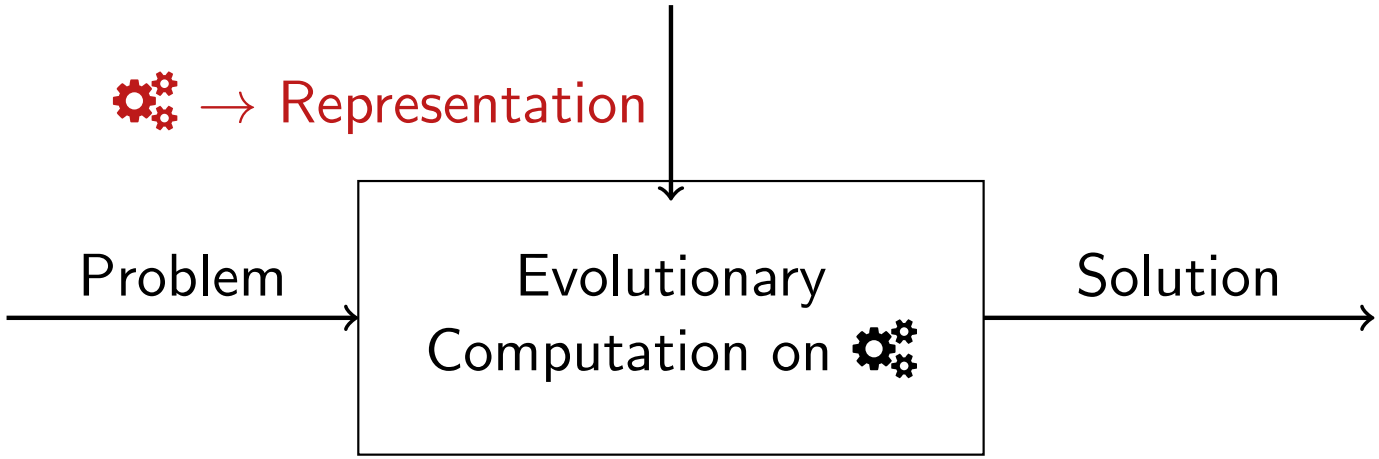
- Individual representation is a key component of every EA
- Humans (EC researchers) put effort in designing good representations

Individual representation



- Individual representation is a key component of every EA
- Humans (EC researchers) put effort in designing good representations
- Can they be designed automatically?

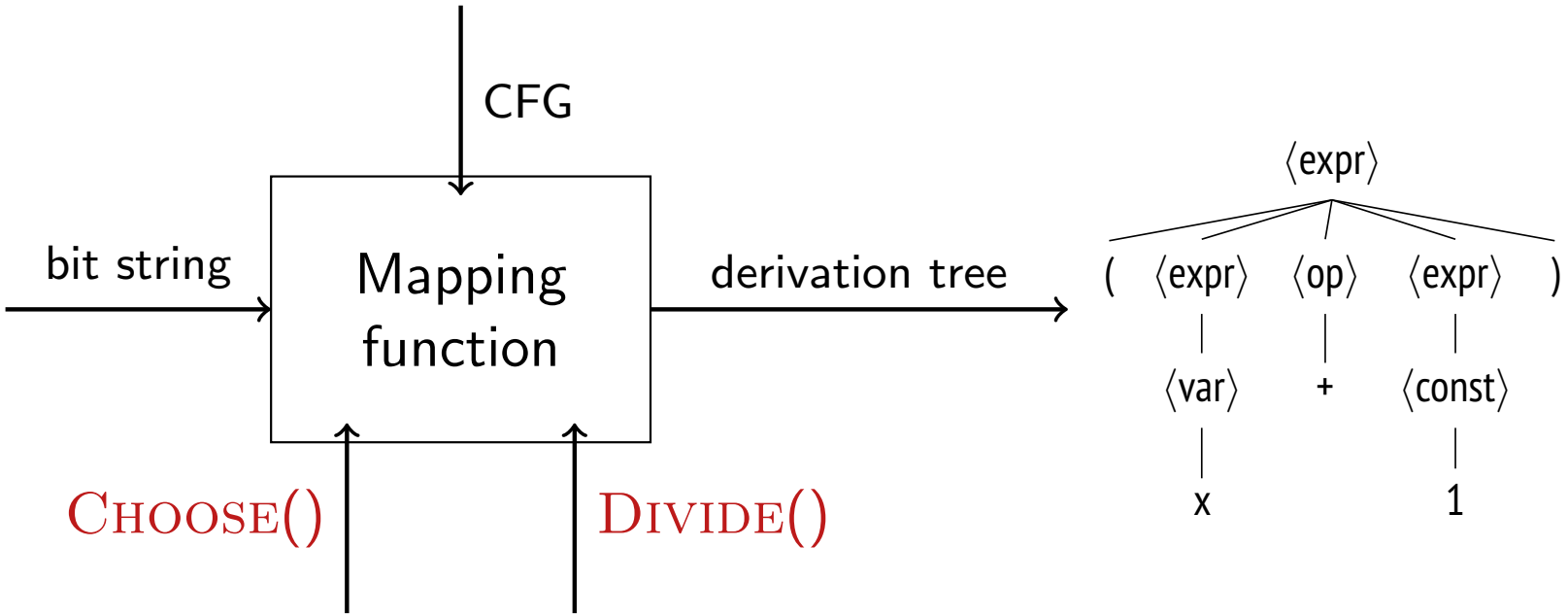
Individual representation



- Individual representation is a key component of every EA
- Humans (EC researchers) put effort in designing good representations
- Can they be designed automatically?

TL;DR: yes, with GP! and they are human-competitive!

The representation of a representation

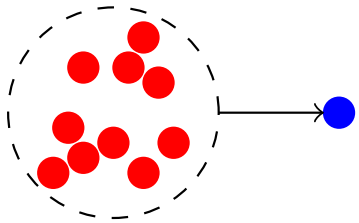


- Modular mapping function which always returns a derivation tree
- Search space of CHOOSE() and DIVIDE() defined by a CFG
- Can express existing representations: GE, HGE, WHGE

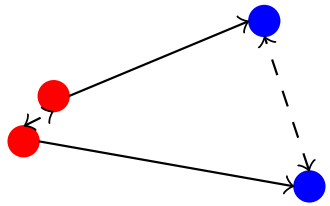
Fitness function

Goal: evolving a representation with good properties

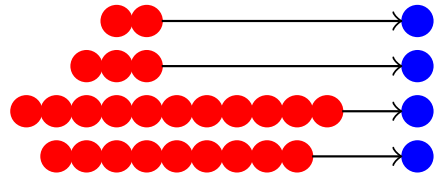
Redundancy (R)



Non-locality (NL)



Non-uniformity (NU)



- "Known" to be important: the lower, the better
- Three variants for reaching this goal: R, R+NL, R+NL+NU

Table of Contents

Experiments

RQ1 Do the evolved representations exhibit better properties than the existing, human-designed ones?

RQ2 Are the evolved representations also more effective when used inside an actual EA?

Experiments

RQ1 Do the evolved representations exhibit better properties than the existing, human-designed ones?

RQ2 Are the evolved representations also more effective when used inside an actual EA?

1. Evolve many representations: fitness as the properties on a set of 3 on 4 problems (*learning*)
2. Choose the most effective: best average final fitness when used in an EA applied to the 4 problems (*validation*)
3. Assess chosen representation also on other 4 problems, not used in learning nor validation (*test*)
   - Comparison against human-designed baselines: GE, HGE, WHGE

RQ1: better in properties

	Learning			Validation		
	R	NL	NU	R	NL	NU
R	0			0.242	0.719	0.311
R+NL	0.03	0.495		0.225	0.606	0.451
R+NL+NU	0.009	0.567	0.032	0.156	0.698	0.214
GE				0.993	1	0.632
GE_{opt}				0.911	0.561	2.036
HGE				0.658	0.572	2.515
WHGE				0.573	0.585	2.814

- On average, lower redundancy and non-uniformity than human-designed!

RQ2: better in search effectiveness

Problem-wise and average percentile rank of the final fitness

	Keijzer6	KLand.-5	KLand.-7	MOPM-3	Nguyen7	Pagie1	Parity-3	Text	Avg.
R	0.077	0.111	0.045	0.066	0.179	0.085	0	0.022	0.075
R+NL	0.04	0.005	0.073	0.017	0.13	0.169	0	0.037	0.061
R+NL+NU	0.106	0.152	0.111	0.025	0.156	0.032	0	0.015	0.075
GE	0.441	0.997	0.997	0.294	0.705	0.637	0.987	0.123	0.647
GE_{opt}	0.07	0.89	0.895	0.015	0.099	0.194	0	0.037	0.282
HGE	0.095	0.147	0.031	0.09	0.29	0.31	0	0.006	0.131
WHGE	0.047	0.147	0.013	0.041	0.094	0.145	0.051	0.01	0.069

- Best evolved representation is better than all the human-designed ones!

Table of Contents

Fundamental problem in EA design

We faced a fundamental, long-standing problem:

- "perhaps the most difficult and least understood area of EA design is that of adapting its internal representation."[1] (2007)
- "How should the representations that are used in evolutionary algorithms, on which variation and selection act, be chosen and justified?"[2] (2017)

[1] De Jong, "Parameter setting in EAs: a 30 year perspective", 2007.

[2] Spector, "Introduction to the peer commentary special section on "On the Mapping of Genotype to Phenotype in Evolutionary Algorithms" by Peter A. Whigham, Grant Dick, and James Maclaurin", Sept. 2017.

Fundamental problem in a broader sense

Our contribution broadens the scope of human-competitive:

- from "solving a specific problem"...
- ...to "designing the overall solution framework" (partially automating the modelling phase)

A challenging scenario as well

Grammatical Evolution:

- great practical interest: works on any CFG-based problem
- non-trivial indirect representation: attracted many studies for a long time
  - experimental studies on properties (R, NL, NU)
  - carefully designed representation variants: GE, πGE, HGE/WHGE (and SGE)

Thanks!