CISUC
CENTRO DE INFORMÁTICA E SISTEMAS DA UNIVERSIDADE DE COIMBRA

THE INFLUENCE OF PROBABILISTIC GRAMMARS ON EVOLUTION

Analysis of Structured Grammatical Evolution Methods

Jessica Mégane · jessicac@dei.uc.pt
Nuno Lourenço · naml@dei.uc.pt
Penousal Machado · machado@dei.uc.pt
Dirk Schweim · schweim@hs-mainz.de

www.cisuc.pt

Context

Genetic programming has been extensively used for program evolution, but the quest for improved algorithms continues. While performance has been a primary focus, researchers have discovered other critical metrics by analyzing algorithm behavior. Issues such as **bloat** (uncontrolled program growth) and **bias towards smaller trees** [1] have been attributed to representation and mapping mechanisms, necessitating thorough analysis of the evolving population.

E), known for **low locality** and **high redundancy**, prompted the introduction of Structured GE [3], followed by Probabilistic SGE (PSGE) [4] and Co-evolutionary PSGE (Co-PSGE) [5], which employed a Probabilistic Context-Free Grammar to bias search through production rule probabilities.These methods demonstrated advancements in performance and locality compared to SGE, highlighting the importance of studying algorithm behavior during evolution for future enhancements.

Probabilistic Context-Free Grammar (PCFG)

Defines the space of syntactically valid solutions. A PCFG has a probability associated with each production rule.

<expr> ::= <expr><op><expr>	[0.00; 0.37]
| <var>	]0.37; 1.00]
<op> ::= +	[0.00; 0.22]
| −	]0.22; 0.39]
| *	]0.39; 0.68]
| /	]0.68; 1.00]
<var> ::= x	[0.00; 0.41]
| y	]0.41; 0.67]
| 1.0	]0.67; 1.00]

Fig.1 Example of PCFG

Structured Grammatical Evolution (SGE)

The genotype is a list of dynamic lists. Each list corresponds to a non-terminal and the elements are ordered with the indexes of the rules to expand.

< expr >	< op >	< var >
[0, 1, 1]	[1]	[2, 0]

Fig.2 Example of the genotype of SGE

Probabilistic SGE (PSGE)

The genotype elements are floats, which represent the probability of selecting a production rule.

<expr>	<op>	<var>
[0.19,0.46,0.87]	[0.27]	[0.32, 0.64]

Fig.3 Example of the genotype of PSGE and Co-PSGE

In the mapping, it is verified whether the codon belongs to the probability range of each production rule of the non-terminal to be expanded and when this condition is verified, the rule is chosen.

<expr> → <expr><op><expr>	(0.19)
<expr><op><expr> → <var><op><expr>	(0.46)
<var><op><expr> → x <op><expr>	(0.32)
x<op><expr> → x - <expr>	(0.27)
x - <expr> → x - <var>	(0.87)
x - <var> → x - y	(0.64)

Phenotype: $x - y$

Fig.4 Example the mapping mechanism of PSGE and Co-PSGE

Co-evolutionary PSGE (Co-PSGE)

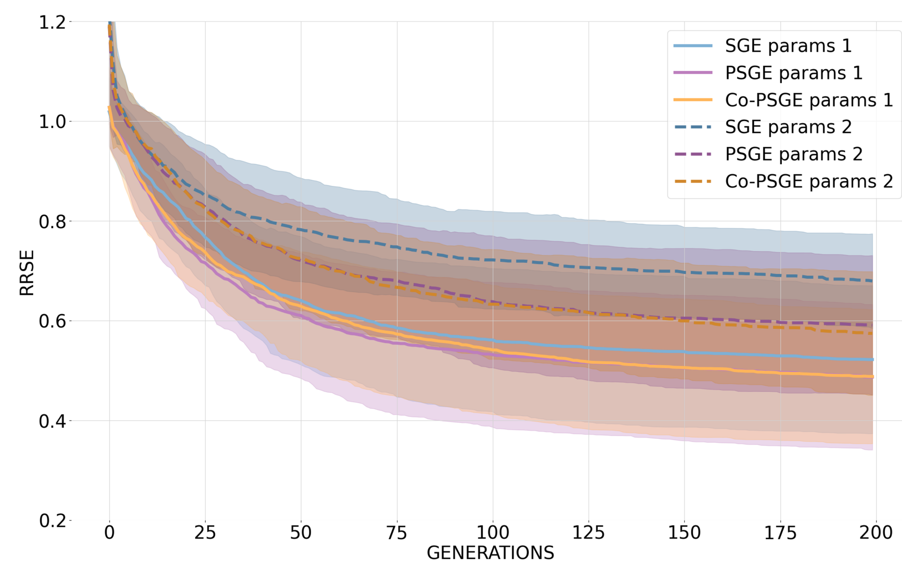
Each individual presents a genotype similar to PSGE, and a PCFG. Each generation the individual can suffer mutation and crossover, but also its grammar can suffer a Gaussian mutation on its probabilities. At most one mutation per non-terminal.

Results analysis

PERFORMANCE

PSGE and Co-PSGE are better than SGE using Params 2:

- Probabilistic approaches have better results with a smaller population
- SGE benefits from a larger sample size



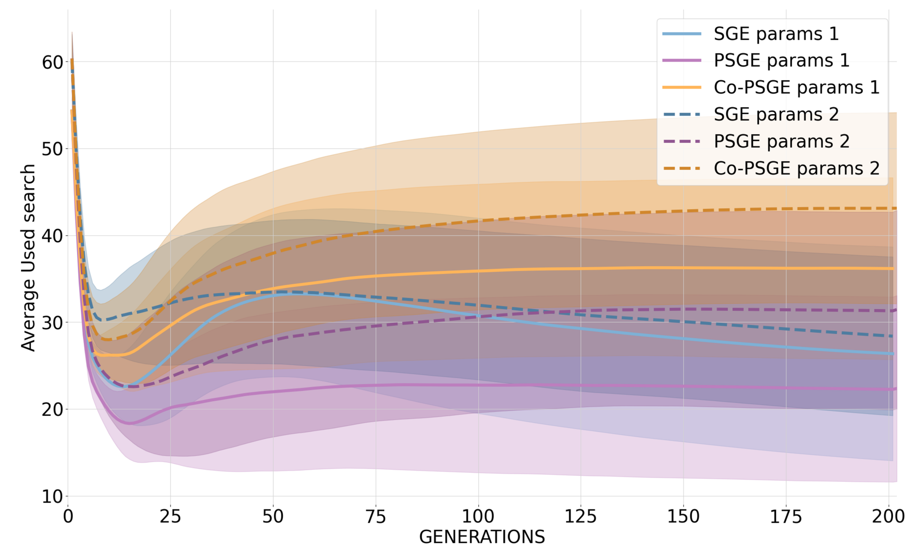
PERCENTAGE OF UNIQUE SOLUTIONS

Algorithms show a higher percentage of unique solutions with a smaller population (Params 2)

- In SGE, the difference between the percentages is lower
- Supports the theory that SGE benefits from a larger sample size

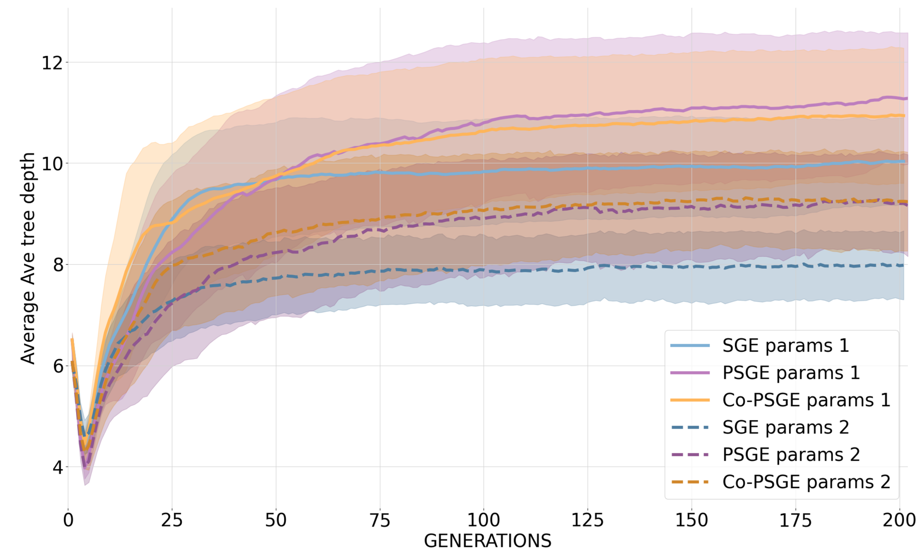
The percentage in SGE decreases

- PSGE starts with lower value, but surpasses SGE



TREE DEPTH

SGE individuals are smaller, and stagnate near the maximum defined

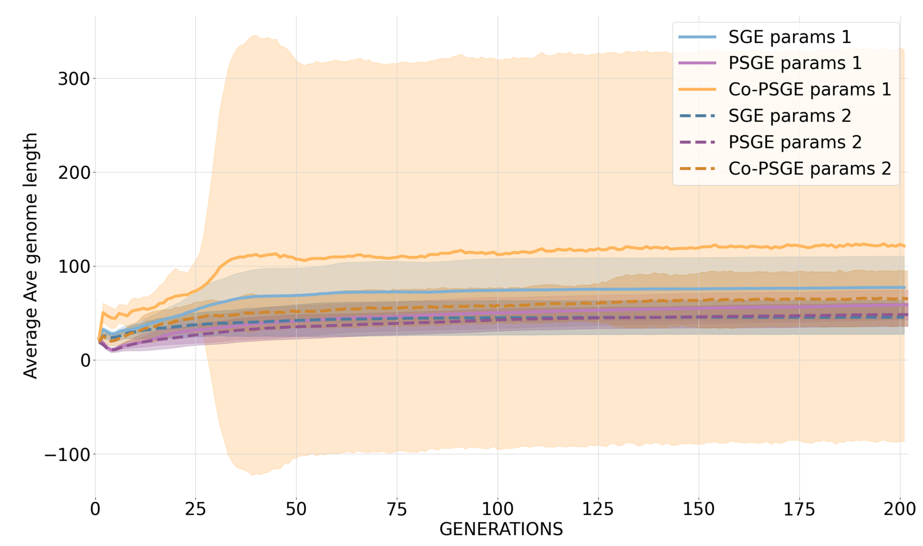


GENOTYPE LENGTH

SGE individuals are smaller, and stagnate near the maximum defined

Co-PSGE has a sudden increase in the average and standard deviation, around 25 generations for Params 1, and 125 generations for Params 2.

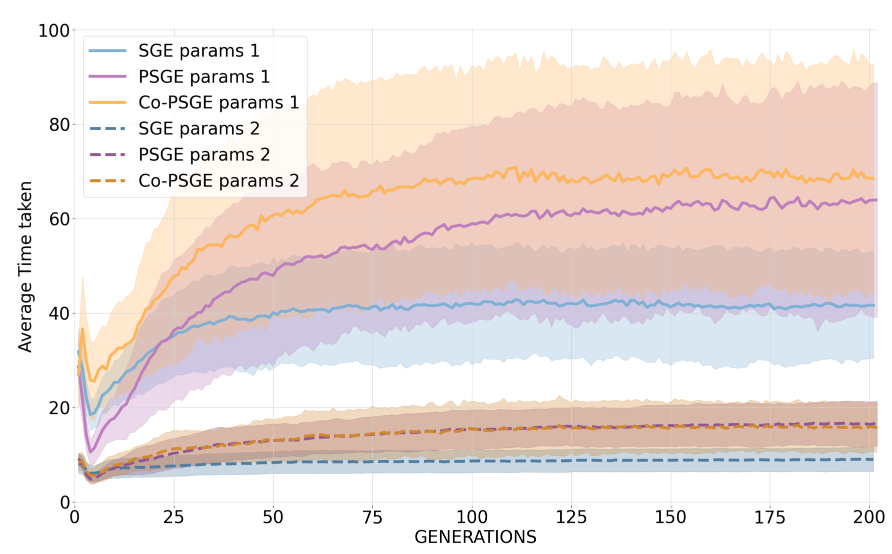
- The algorithm might suffer from bloat, as there is no corresponding improvement in fitness



EXECUTION TIME

Probabilistic approaches take longer

With a bigger population, Co-PSGE takes more time than PSGE



Experimental Setup

Experiments set for finging the **pagie polynomial**, evaluated using the root relative squared error.

Executed over **100** runs with two sets of parameters.

	Params 1	**Params 2**
Population Size	1000	250
Generations	200	
Elitism Count	100	25
Mutation Rate	0.05	0.10
Crossover Rate	0.90	0.90
Tournament	3	
Max Depth	10	8

	p-value	effect size
Params 1		
SGE - PSGE	0.225	~
SGE - Co-PSGE	0.099	~
PSGE - Co-PSGE	1.355	~
Params 2		
SGE - PSGE	0.00	++
SGE - Co-PSGE	0.00	++
PSGE - Co-PSGE	0.371	~

Fig.11 Results of the Mann-Whitney post-hoc statistical tests applied to the performance data.
The Bonferroni correction is used, with significance level of 0.05

Future Work

- Repeat analysis with different parameters (mutation and crossover rate, population size, and depth)
- Compare results in different problems
- Perform study with different grammars
- Compare with other grammar-based methods, for example CFG-GP and GE

Paper PDF and source code

References

[1] Dirk Schweim and Franz Rothlauf. 2018. An Analysis of the Bias of Variation Operators of Estimation of Distribution Programming. In Proceedings of the 800 Genetic and Evolutionary Computation Conference (Kyoto, Japan) (GECCO '18). 801 Association for Computing Machinery, New York, NY, USA, 1191–1198
[2] C. Ryan, M. O'Neill, and J.J. Collins (Eds.). 2018. Handbook of Grammatical Evolution. Springer International Publishing.
[3] N. Lourenço, F. B. Pereira, and E. Costa. 2016. SGE: A Structured Representation for Grammatical Evolution. In Lecture Notes in Computer Science. Springer International Publishing, 136-148
[4] J. Megane, N. Lourenço, and P. Machado. 2022. Probabilistic Structured Grammatical Evolution. In 2022 IEEE Congress on Evolutionary Computation (CEC). IEEE.
[5] J. Mégane, N. Lourenço, and P. Machado. 2022. Co-evolutionary probabilistic structured grammatical evolution. In Proceedings of the Genetic and Evolutionary 776 Computation Conference. ACM

This work was funded by FEDER funds through the Operational Programme Competitiveness Factors - COMPETE and national funds by FCT - Foundation for Science and Technology (POCI-01-0145-FEDER-029297, CISUC - UID/CEC/00326/2020) and within the scope of the project A4A: Audiology for All (CENTRO-01-0247-FEDER-047083) financed by the Operational Programme for Competitiveness and Internationalisation of PORTUGAL 2020 through the European Regional Development Fund. The first author is funded by FCT, Portugal, under the grant 2022.10174.BD.

1290 UNIVERSIDADE D COIMBRA · fct Fundação para a Ciência e a Tecnologia · LASI Intelligent Systems Associate Laboratory · Cofinanciado por: COMPETE 2020 · PORTUGAL 2020 · UNIÃO EUROPEIA Fundos Europeus Estruturais e de Investimento