Cartesian Genetic Programming: Why No Bloat?

Andrew James Turner and Julian Francis Miller

andrew.turner, julian.miller @york.ac.uk

THE UNIVERSITY of York
DEPARTMENT OF ELECTRONICS

Abstract

For many years now it has been known that Cartesian Genetic Programming (CGP) does not exhibit program bloat. Two possible explanations have been proposed in the literature: neutral genetic drift and length bias. This paper empirically disproves both of these and thus, reopens the question as to why CGP does not suffer from bloat.

Measuring Bloat

Bloat can be formally measured using the equation below; adapted form [3].

$$B(g) = \frac{\dfrac{\hat{A}(g) - \bar{A}(0)}{\bar{A}(0)}}{\dfrac{\bar{F}(0) - \hat{F}(g)}{\bar{F}(0)}}$$

- $B(g)$ – bloat at generation g
- $\hat{A}(g)$ – program size of fittest chromosome
- $\bar{A}(g)$ – average program size
- $\bar{F}(g)$ – average population fitness
- $\hat{F}(g)$ – fitness of fittest chromosome

This equation gives the ratio of the increase in program size to the increase in fitness. If program size increases disproportionally to fitness then bloat is present.

Previous theories why CGP does not bloat

Neutral Genetic Drift [1]

1. When trapped in local optima, only mutations which do not decrease fitness are preserved.
2. Given a chromosome, there are more equally fit chromosomes which are larger than smaller.
3. Therefore there is a pressure to bloat.
4. However, CGP chromosomes contain inactive genes.
5. Mutations to inactive genes do not affect fitness or program size.
6. Therefore mutations to inactive genes are more likely to be preserved than mutations to active genes.
7. Therefore bloating does not occur in CGP.

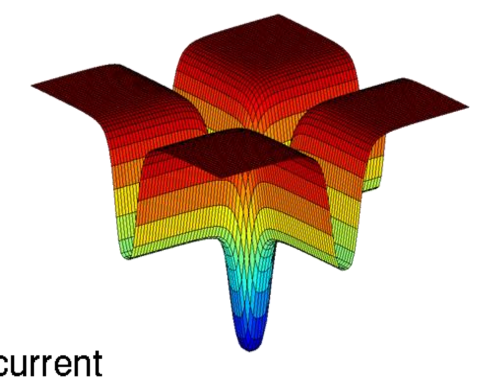
Length Bias [2]

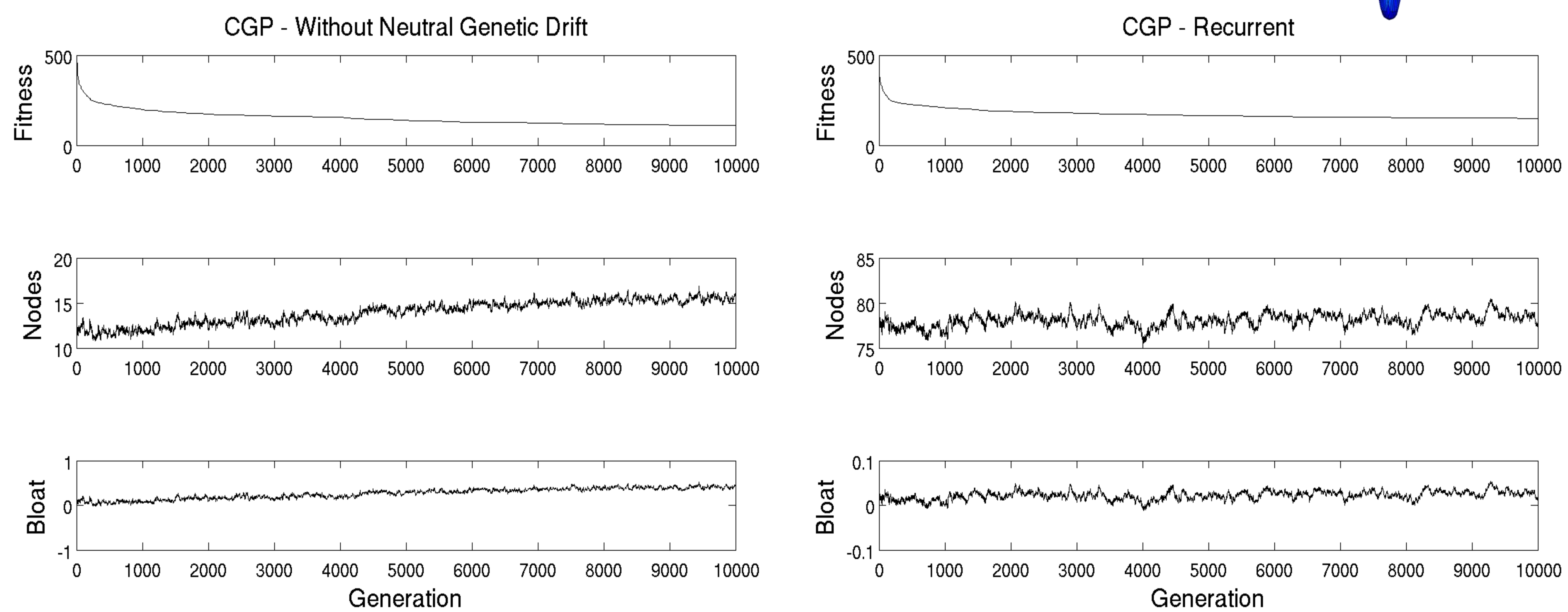
1. CGP typically evolves acyclic programs.
2. Here, nodes positioned closer to the inputs can be connected to by more nodes than those closer to the outputs.
3. Therefore nodes closer to the inputs have a higher probability of being active.
4. This produces a length bias towards smaller program sizes, preventing bloat.

Experiments

Both of these theories can be interrogated by removing the underlying assumption they each make. Neutral genetic drift can be prevented by disallowing mutations to inactive genes. Length bias can be prevented by allowing recurrent / cyclic connections. The experiments presented here are applied to the Pagie1 symbolic regression benchmark:

$$f(x_1, x_2) = \frac{1}{1 + x_1^{-4}} + \frac{1}{1 + x_2^{-4}}$$





Conclusion

As can be seen in the above figures, neither preventing neutral genetic drift or allowing recurrent connections causes bloat in CGP. Therefore, nether neutral genetic drift or length bias is the cause of CGP not bloating. Although no new explanations are proposed, this research reopens the topic for further investigation.

References

[1] Miller, J.F.: What bloat? Cartesian Genetic Programming on Boolean problems. In: GECCO-2001, ACM (2001) 295–302
[2] Goldman, B.W., Punch, W.F.: Length bias and search limitations in Cartesian Genetic Programming. In: GECCO-2013, ACM (2013) 933–940
[3] Vanneschi, L., Castelli, M., Silva, S.: Measuring bloat, overfitting and functional complexity in genetic programming. In: GECCO-2010, ACM (2010) 877–884