

Recombination Hotspots Promote the Evolvability of Modular Systems

Ari Larson
University of Vermont
ariel.larson@uvm.edu

Ken Livingston
Vassar College
livingst@vassar.edu

Jodi Schwarz
Vassar College
joschwarz@vassar.edu

Anton Bernatskiy
University of Vermont
anton.bernatskiy@uvm.edu

Nicholas Livingston
Vassar College
nilivingston@vassar.edu

Marc Smith
Vassar College
mlsmith@vassar.edu

Collin Cappelle
University of Vermont
collin.cappelle@uvm.edu

John Long
Vassar College
jolong@vassar.edu

Josh C. Bongard
University of Vermont
josh.bongard@uvm.edu

ABSTRACT

Random recombination in evolutionary algorithms can be counterproductive in systems that evolve increasing modularity, because such operators do not preserve community structures during their development. Partly because of this, methods have been proposed that derandomize recombination by placing potential crossover locations under evolutionary control. Since crossover is likely to be particularly useful when genetic material that generates incipient phenotype modules is recombined, there may be an advantage to seeking such modularity directly in the phenotype and probabilistically focusing recombination at such "hotspot" locations. Here we show that such phenotypically-aware crossover operators can outcompete random or evolved crossover points as the size of the system being evolved grows. As this crossover operator can be viewed as epigenetic, and as epigenetic processes seem to be common in biological systems, other such epigenetic mechanisms may further improve future evolutionary algorithms.

Keywords

Recombination Operators, Crossover, Modularity

1. INTRODUCTION

Modularity is a highly conserved mode for organizing biological processes across all domains of life, and in evolving systems it can increase evolvability by allowing modification to one module without disrupting existing functionality in other modules (e.g. [4]). Modular systems also allow for the combination of useful modules from different systems, potentially increasing the usefulness of genetic recombination, as recombination is most effective when functionally indepen-

dent modules are crossed between individuals. We propose that this relationship between modularity and recombination can be utilized by directly referencing the modularity of a network's phenotypic structure during recombination, using the Q metric for modularity [3]. We introduce the concept of modularity-guided linkage learning, which uses network phenotypic structure to guide recombination probabilistically.

2. METHODS

Here we employ an evolutionary algorithm introduced by [1] for evolving random boolean networks that, when coupled with an appropriate fitness function and a bias toward sparse networks, was shown to evolve increasingly fit and modular Boolean networks without explicitly selecting for modularity [1]. To this basic algorithm we add one of five crossover strategies, and compare the relative performances of the five resulting experiment variants E_1 through E_5 .

In experimental set up E_1 we evolve modular networks using mutation only, with no crossover, following [1] as closely as possible. Only a few modifications were made to improve the running time efficiency of the algorithm. In E_2 , we add a random crossover strategy. In E_3 , we limit crossover to the known point of emerging modularity for fit networks in this problem. In E_4 , we evolve crossover probability distributions in parallel with individual networks, in the style of linkage learning (e.g. [2]). Each evolving network is thus associated with an additional evolving vector which probabilistically determines crossover locations for that network. For the sake of simplicity, we limit crossover to n possible partitions for an $n \times n$ network, corresponding to splitting the adjacency matrix of the network horizontally along an $i \in [1, \dots, n-1]$. When two adjacency matrices A_1 and A_2 have been selected for crossover, their probability distributions are added, the resulting vector is re-normalized to sum to one, and the crossover index i is chosen from this probability distribution. Two child matrices C_1 and C_2 are produced as $A_2[i : n-1, :]$ appended to $A_1[0 : i-1, :]$ and $A_1[i : n-1, :]$ appended to $A_2[0 : i-1, :]$, respectively. Finally, in E_5 , the phenotypic structure of the network is used to guide crossover using the Q metric for modularity. The normalized modularity ratings for the $n-1$ network partitions $P_1 = \{p_1 = \{n_1\}, p_2 =$

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

GECCO'16 Companion July 20-24, 2016, Denver, CO, USA

© 2016 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-4323-7/16/07.

DOI: <http://dx.doi.org/10.1145/2908961.2908990>

Table 1: Relative search abilities and modularity scores of the experiments. The comparison $e(E_x) < e(E_y)?$ indicates a test to determine whether the error of experiment x is significantly lower than experiment y . The p values resulting from this comparison are reported in column 6. The comparison $Q(E_x) < Q(E_y)?$ indicates a test to determine whether the modularity of experiment x is significantly higher than experiment y . The p values resulting from this comparison are reported in column 9. Asterisks indicate statistical significance.

	1	2	3	4	5	6	7	8	9
A	Crossover Style	Nodes	Trials	Comparison	Mean Error	p	Comparison	Mean Q	p
B	E ₁	10	60		0.037			0.161	
C	E ₂	10	60	$e(E_2) < e(E_1)?$	0.026	0.08063	$Q(E_2) > Q(E_1)?$	0.176	0.42108
D	E ₃	10	60	$e(E_3) < e(E_2)?$	0.009	0.00102*	$Q(E_3) > Q(E_2)?$	0.252	0.00039*
E	E ₄	10	60	$e(E_4) < e(E_2)?$	0.015	0.01974*	$Q(E_4) > Q(E_2)?$	0.265	0.00005*
F	E ₅	10	60	$e(E_5) < e(E_2)?$	0.011	0.00039*	$Q(E_5) > Q(E_2)?$	0.280	0.00001*
F	E ₅	10	60	$e(E_5) < e(E_4)?$	0.011	0.18582	$Q(E_5) > Q(E_4)?$	0.280	0.15277
G	E ₄	12	20		0.028			0.316	
H	E ₅	12	20	$e(E_5) < e(E_4)?$	0.017	0.02541*	$Q(E_5) > Q(E_4)?$	0.274	0.21429

$\{n_2, \dots, n_{10}\}$, $P_2 = \{p_1 = \{n_1, n_2\}, p_2 = \{n_3, \dots, n_{10}\}\}$, \dots , $P_{n-1} = \{p_1 = \{n_1, \dots, n_9\}, p_2 = \{n_{10}\}\}$ are computed, and the resulting $(n-1)$ -element vectors are employed for guiding crossover as described for E_4 .

3. RESULTS AND DISCUSSION

Table 1 reports the relative performance of the five experiments and the relative modularity of the networks produced by these experiments. We reproduced Espinosa-Soto and Wagner’s results for emergent modularity for this problem. Introducing random crossover (E_2) did not significantly improve evolvability at this sample size (Table 1, cell C6). Using fixed crossover (E_3) significantly improved evolvability, with an average difference of 0.017 from random crossover (Table 1, cell D6). Using evolved crossover hotspots (E_4) significantly improved evolvability, with an average difference of 0.011 from random crossover (Table 1, cell E6). Using modularity-based crossover hotspots (E_5) did significantly improve evolvability, with an average difference of 0.015 from random crossover (Table 1, cell F6). In larger networks of 12 nodes, using evolved and modularity-based crossover strategies, we found that while both strategies produced similarly modular networks, the modularity-based strategy (E_5) produced networks of higher fitness in the same number of generations (Table 1, cell H6). This difference was not observed in networks of 10 nodes.

E_4 and E_5 did not yield networks with significantly different error or modularity when those networks were composed of 10 nodes, but E_5 did yield significantly lower errors compared to E_4 when the two methods were applied to slightly larger 12-node networks (Table 1, cell H6). This suggests that modularity-aware crossover may be more scalable than evolved crossover: indeed the dimensionality of the search space increases at a faster rate for E_4 than for E_5 as the number of nodes in the evolved networks increases.

4. FUTURE WORK

Our work suggests that phenotypic crossover mechanisms may be fruitful avenues of further research. For the problem investigated, we found that a modularity-based phenotypic crossover strategy yielded networks with lower error in the same number of generations than a similar approach using random crossover. In addition, this strategy scaled to larger

networks better than a similar strategy that used an evolved crossover approach.

Phenotypic crossover mechanisms are useful in this problem, but the model used is somewhat contrived to yield solutions with two highly defined modules [1]. In future work, we plan to investigate the efficacy of phenotypic recombination for problem domains in which it is not known whether, how much, or where modularity may be useful.

We have demonstrated that attention to phenotypic structure using the Q metric can be helpful in evolutionary algorithms. However, Q can be costly to compute, especially in larger networks. In future work, we would like to assess whether the evolutionary advantages granted by phenotypically-aware crossover outweigh the computational costs associated with modularity estimation. Additionally, we would like to use Q to guide the development of crossover locations periodically for a population, without relying on it at every crossover occurrence for each individual.

Acknowledgements

This work was supported by the National Science Foundation (NSF) under projects PECASE-0953837 and INSPIRE-1344227

5. REFERENCES

- [1] C. Espinosa-Soto and A. Wagner. Specialization can drive the evolution of modularity. *PLoS Comput Biol*, 6(3):e1000719, 2010.
- [2] D. G. Georges Harik. Learning gene linkage to efficiently solve problems of bounded difficulty using genetic algorithms. *FOGA*, 4:247–262, August 1996.
- [3] M. Newman. Modularity and community structure in networks. *Proc Natl Acad Sci USA*, 103(23):8577–8582, April 2006.
- [4] R. A. Watson. *Compositional evolution: the impact of sex, symbiosis and modularity on the gradualist framework of evolution*. Mit Press, 2006.