A Computational Model of Symbiotic Composition in Evolutionary Transitions

Richard A. Watson & Jordan B. Pollack

Biosystems 2002

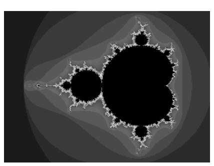
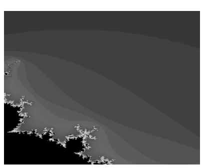
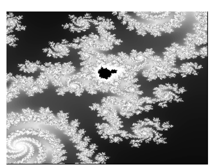
- *Major Evolutionary Transitions*: Genes → Gene Networks → Bacteria → Eukaryotes → Multicellular Organisms → Societies
- All seem to involve *composition* of smaller, simpler entities into larger, more complex ones
- Contrast with *accretion* of traits by classical Darwinism / GA's.

Related Ideas

- *Modularity*: Identification of meaningful components that can be re-used to make subequent variation more "informed"
- *Division of Labor*: Decomposition of a complex adaptation into simpler adaptations that each can be evolved by semi-independent processes
- *Divide-and-Conquer*: Algorithmic approach in which different parts of problem are solved separately, and separate solutions are then combined.

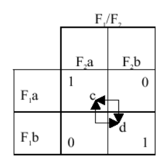
SEAM Model and Issues

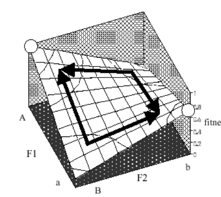
- SEAM: *S*ymbiotic *E*volutionary *A*daptive *M*odel (provides "seams" between combined elements)
- *Fractal Fitness Landscape*: Ruggedness at all scales (never get to a smooth interval): *Scale-Invariance*

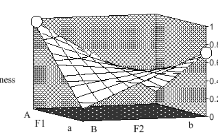




SEAM Model and Issues

- *Epistasis*: Value of change in one bit is affected by value of other bit.
- Can result in *fitness saddles*:

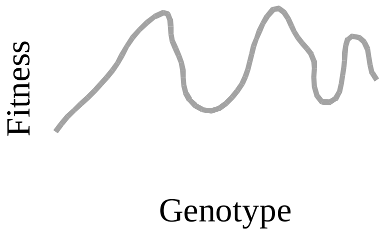






Related Issues

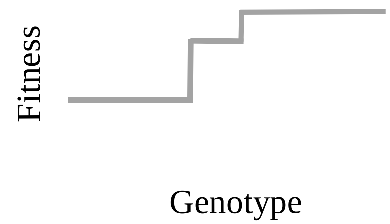
Rugged Fitness Landscape concept suggests prevalence of local optima in fitness function.



However...

Related Issues

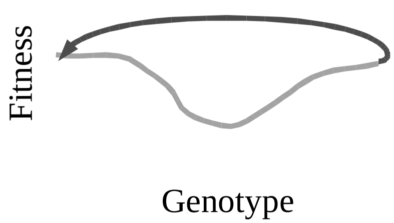
• *Neutral* (Not Neural!) *Networks*: Existence of fitness landscapes where variation is neutral (no effect on fitness) until sudden change:



Related Issues

• *Exaptation*: Features adapted for one purpose are “co-opted” for another, introducing a large phenotype change.

• *Extra-Dimensional Bypass*: Adding another dimension to the genotype can provide a way around a “valley”:



Related Issues

• *Genetic Linkage*: Funciontally related bits (e.g., schema, epistasis) must stay together under genetic operations (crossover).

Modularity and Credit Assignment

- Modularity is expliclitly encoded by some types of models we have studied: Messy GA (Goldberg); GP (Koza)
- *Credit assignment problem*: How to apportion fitness to individual modules?
- Manifests itself in “parasites” and “hitchiking”

The Composition Model (SEAM)

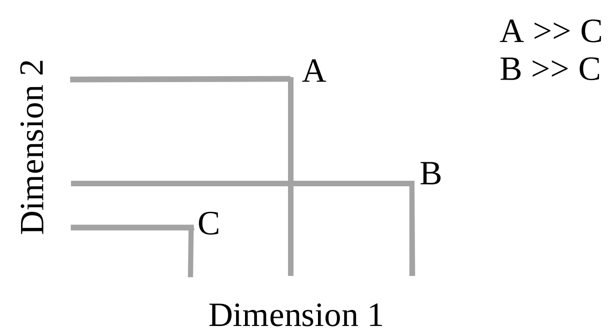
- No genotype/phenotype distinction
- Set of entities represents “ecosystem” of different species.
- No crossover or mutation; instead, use symbiotic joining (union) of randomly-picked entities.
- Then test whether joined “super-entity” is fitter than either parent.
- Fitness of entity changes in different environments, made of other individuals in ecosystem.

The Composition Model (SEAM)

- *Multi-objective optimization*: Fitness function has several dimensions (c.f. report card: math, English, history, ...)
- *Pareto Dominance*: Individual A dominates B iff B is no better than A along any dimension, and A is better than B along at least one:

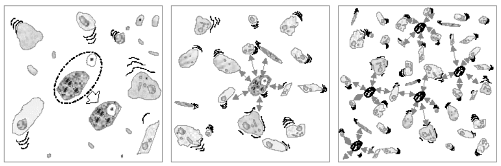
The Composition Model (SEAM)

Pareto Dominance



The Composition Model (SEAM)

- Ecosystem intialized with many different small entities.
- Pairs of entities are picked at random to see whether they form a stable join (better than either by itself).
- If so, they form a new entity; otherwise, they are returned to the population.



a) New entities are created by joining two existing entities together. b) The fitness of an entity is dependent on its environmental context. c) An entity is placed in many contexts to test the stability of a new join.

Entities and their Composition

- A species is a set of entities with identical feature values (F1 = 0, F2 = 1, F5 = 1, etc.)
- Composite is created from joining entities of different species.
- So different entities should also be able to specify different *sets* of features:

E1 = [F1=0, F3=1, F6=1] +

E2 = [F2=0, F3=0, F5=1, F7=0] =

E3= [F1=0, F2=0, F3=1, F5=1, F6=1, F7=0]

Pareto Dominance to Determine Whether a Composition is Preferred

- Each entity is evaluated in a set of *contexts* (fully-specified feature values):

```
---0-11---110---
0110101100010011
0110111100110011
```

---0-11---110--- $\boldsymbol{x}$, an entity specifies a partial set of feature values.
0110101100010011 θ, an 'environmental context' is a complete set of feature values.
0110111100110011 $S(x,\theta)$, the entity x superimposed on the context θ.

- Overall fitness of entity p is weighted sum of context-sensitive fitnesses over all contexts :

$$F(p) = \sum_{\theta \in Contexts} \left(\lambda_{(\theta,p)} csf(p,\theta) \right)$$

Pareto Dominance to Determine Whether a Composition is Preferred

- But generally don't know λ's
- So use Pareto dominance:

$$x >> y \Leftrightarrow (\forall\theta : csf(x,\theta) \geq csf(y,\theta) \ \text{AND} \ \exists\theta : csf(x,\theta) > csf(y,\theta).$$

- Then a combination $a+b$ is stable iff $a+b >> a$ and $a+b >> b$.

Building Environmental Contexts

- Recall that environmental contexts θ are formed from other entities in ecosystem.
- But other entities are also partially specified.
- So we randomly select entities until a complete context is formed:

```
                 a:   --0---1-
                 b:   01------
                 c:   -0---0--
                 d:   ----1-0-
                 e:   ------10
                 f:   ---0-00-
Resultant context     01001010
```

Summary: The SEAM Algorithm

- Variable size entities and a variation operator based on composition
- Building environmental contexts from other co-adapting entities in the ecosystem
- Testing (in)stability of compositions by testing for Pareto dominance of the composition over the component entities

Summary: The SEAM Algorithm

- Initialise ecosystem, E, to random, single-feature, entities.[1]
- Repeat until *stopping condition*:
  - Remove two entities at random from the ecosystem → ***a*** & ***b***.
  - Produce ***a+b***=***S***(***a***,***b***), using composition (see Eq.1).
  - If ***unstable***(***a+b***, ***a***, ***b***) return ***a*** and ***b*** to ecosystem, else add ***a+b*** to ecosystem.

where ***unstable***(***a+b***, ***a***, ***b***) ⇔

$$\exists\theta \in \textbf{\textit{Contexts}}: (f(S(a,\theta)) > f(S(a+b,\theta)) \text{ OR } f(S(b,\theta)) > f(S(a+b,\theta)))$$

where ***Contexts*** is a random set of contexts each built by composing together other members of the current ecosystem, E, using ***S****(E) (see Eqs. 3 & 4).

Comparision of Pareto Dominance with Traditional GA Selection

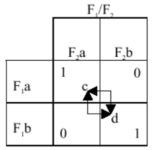
- Recall that in Pareto dominance we keep "non-dominated" (good along some dimension) individuals.
- Tradition GA collapses all dimensions into one, and selects better individuals on that single dimension.
- This may cause GA to converge to minor variants of a "best on average" individual – poor solution.
- SEAM avoids early convergence by using Pareto dominance.

Comparision of SEAM with Messy GA

- Like Messy GA, SEAM uses partially-specified individuals.
- Messy GA fills in unspecified positions using hill-climbed “templates”
- SEAM fills in using other entities in ecosystem.
- This helps SEAM reduce the number of contexts it needs to sample.
- So using other entities is an important feature of *scalability* in SEAM.

A Scale-Invariant Fitness Landscape

- Recall two-feature epistasis:



- We can expand this to four features:

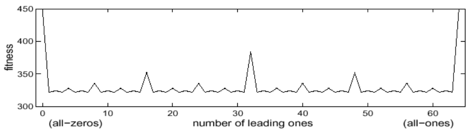
F1	F2	F3	F4	Fitness
0	0	0	0	2
1	1	1	1	2
0	0	1	1	1
1	1	0	0	1
else				0

A Scale-Invariant Fitness Landscape

- Ultimately, we get “HIFF” (Hierarchical Iff):

$$F(B) = \begin{cases} 1, & \text{if } |B| = 1 \\ |B| + F(B_L) + F(B_R), & \text{if } |B|>1 \text{ and } (\forall i{:}b_i=0 \text{ OR } \forall i{:}b_i=1) \\ F(B_L) + F(B_R) & \text{otherwise} \end{cases}$$

- Produces a fractal fitness landscape:



Relation to Evolutionary Games

- Recall Prisoner's Dilemma:

		Player B: Cooperate	Player B: Defect
Player A	Cooperate	3 3	0 5
	Defect	5 0	1 1

- Can view epistasis as this sort of game
- “Deliberately dissolves” distinction between epistasis and evolutionary games

Experiment

Compare SEAM to two other evolutionary algorithms:

1) Random-Mutation Hill-Climbing (flip a bit and keep new string if it's fitter)

2) GA with *deterministic crowding* (diversity maintenance)

Random Mutation Hill-Climbing (RMHC; Mitchell p. 129)

1. Choose a string at random. Call this *best-evaluated*.
2. Choose a locus at random to flip. If the flip leads to an equal or higher fitness, then set *best-evaluated* to the resulting string.
3. Go to step 2 until an optimum string has been found or a maximum number of evaluations has been performed.
4. Return the current value of *best-evaluated.*

Deterministic Crowding

- Initialize population.
- Repeat until stopping condition:
  - Pick two parents, p1 & p2, at random from the population.
  - Produce a pair of offspring, c1 & c2, using recombination, and mutation.
  - Pair-up each offspring with one parent according to the pairing rule below.
  - For each parent/offspring pair, if the offspring is fitter than the parent then replace the parent with the offspring.

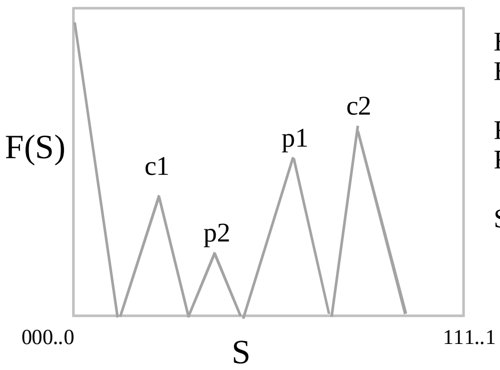
Pairing rule: if H(p1,c1)+H(p2,c2) <H(p1,c2)+H(p2,c1) then pair p1 with c1, and p2 with c2, else pair p1 with c2, and p2 with c1, where H gives the genotypic Hamming distance between two individuals.[4]

Deterministic Crowding: Pairing Rule

- *Hamming Distance*: Total # of different bits
- e.g., H(00, 00) = 0; H(00,01) = 1; H(01,10) = 2
- *Q*: Why compare parent to more-similar (vs. less-similar) child?
- Remember, goal is to maintain diversity!

Deterministic Crowding: Pairing Rule

A: Comparing parent to less-similar child could result in “specialist old fogies” holding back “promising youngsters”:

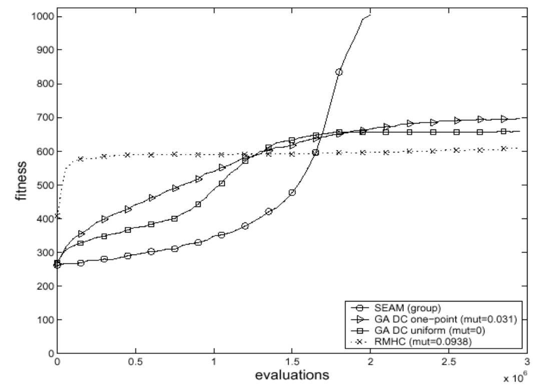


H(c1,p1)+H(c2,p2) > H(c1,p2)+H(c2,p1)

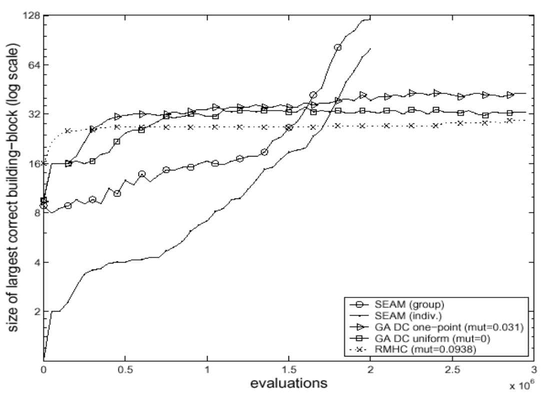
F(c1)>F(p2)
F(c2)>F(p1)

So repalce p's with c's.

Results: Max Fitness



Results: Building-Block Size



Discussion

- SEAM provides a concrete illustration of changing units of variation and of selection.
- New entities selected at one level of organization provide components for entties at next level.
- Variation mechanism scales up with size of entities.
- Composition provides a divide-and-conquer solution to problems with fractal fitness landscape.
- Random variation cannot solve this kind of problem.

Discussion

• So units discovered by SEAM are not just large, but also "usefully informed" by prior adaptation.
• Sets of features exchanged by crossover are abitrary and so do not provided meaningful modules.
• Sets of features exchanged by composition are not arbitrary, and so provide meaningful modules.

Discussion: Baldwin Effect

• Recall Baldwin Effect: Learning (short-term) guides evolution (long-term).
• In SEAM, composition (short-term) guides evolution (long-term).
• In both, *canalization* (fixation) of new variants can occur.

Conclusions

• Random variation is source of innovation, but is "inherently opposed to the heritably of extant complexity."
• *Q*: How to suppress variation without suppressing innovation?
• *A*: Individual entities may be stable, but composition of them may still provide opportunity for innovation.
• Both GA and RMHC have limited ability to cross fitness saddles. SEAM (apparently) does not.