Hyper-heuristics Tutorial

Daniel R. Tauritz (dtauritz@acm.org)
Head of Biomimetic Artificial Intelligence Research Group, Auburn University

John Woodward (J.Woodward@qmul.ac.uk)
Head of Operational Research Group, Queen Mary University of London

https://gecco-2021.sigevo.org/

GECCO '21 Companion, July 10-14, 2021, Lille, France
© 2021 Association for Computing Machinery.
ACM ISBN 978-1-4503-8351-6/21/07…$15.00
https://doi.org/10.1145/3449726.3461418

Instructors

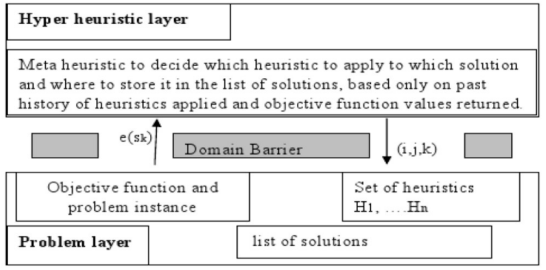
Daniel R. Tauritz is an Associate Professor in the Department of Computer Science and Software Engineering at Auburn University (AU), Interim Director and Chief Cyber AI Strategist of AU's Auburn Cyber Research Center, Head of AU's Biomimetic Artificial Intelligence (BioAI) Research Group, Director of AU's Biomimetic National Security Artificial Intelligence (BONSAI) Laboratory, Guest Scientist at Los Alamos National Laboratory (LANL), and Director of the LANL/AU Cyber Security Sciences Institute. He received his Ph.D. in 2002 from Leiden University. His research interests include the design of hyper-heuristics and self-configuring evolutionary algorithms and the application of computational intelligence techniques in cyber security, critical infrastructure protection, and program understanding.

John R. Woodward is a Lecturer at Queen Mary University of London, and is Head of the Operational Research Group, and previously for four years was a Lecturer with the University of Nottingham and also Stirling. He holds a BSc in Theoretical Physics, an MSc in Cognitive Science and a PhD in Computer Science, all from the University of Birmingham. His research interests include Automated Software Engineering, particularly Search Based Software Engineering, Artificial Intelligence/Machine Learning and in particular Genetic Programming. He has worked in industrial, military, educational and academic settings, and been employed by EDS, CERN and RAF and three UK Universities.

John's perspective of hyper-heuristics

Domain Barrier

$$HH : (H \times \mathbb{N}^2 \times \mathbb{R})^* \rightarrow H \times \mathbb{N}^2$$



Conceptual Overview

Combinatorial problem e.g. Travelling Salesman
Exhaustive search ->heuristic?

Genetic Algorithm
heuristic – permutations

Travelling Salesman

Tour

Single tour NOT EXECUTABLE!!!

Genetic Programming
code fragments in for-loops.

Travelling Salesman Instances

TSP algorithm

EXECUTABLE on MANY INSTANCES!!!

Give a man a fish and he will eat for a **day**.
Teach a man to fish and he will eat for a **lifetime.**

Scalable? General?
New domains for GP

One Man – One/Many Algorithm

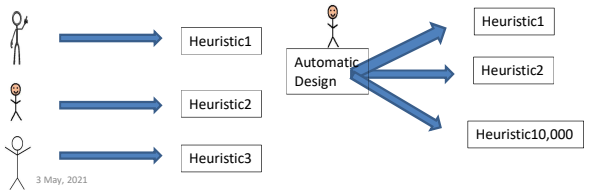
Genetic Programming
{+, -, *, /}
{AND, OR, NOT}

Automatically designed heuristics (this tutorial)

First year university course On Java, as part of a computer Science degree

LARGE Software Engineering Projects

Increasing "complexity"



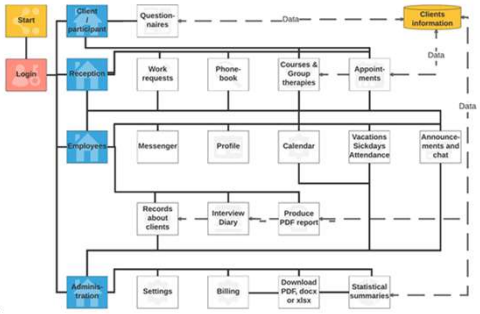
3 May, 2021

Janus Manager

- Management system for rehabilitation
- Stores client info
- A tool for
  - Administration
  - Communication
  - Producing reports
  - Predicting outcomes

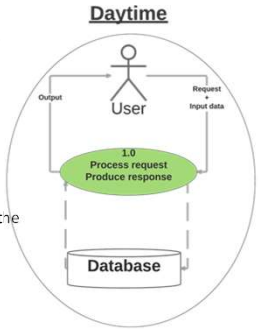
Janus Manager

- ~40 Users
  - Specialists
  - Admin staff
- 1000+ clients
  - ~150 active
- The code
  - Python
  - 25,000+ LOC
  - 600+ functions
  - 300 Classes
  - Run as web service



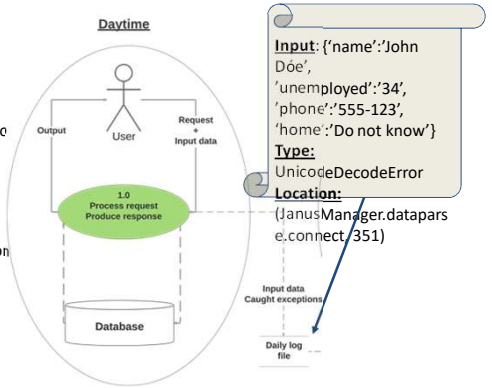
Daily activity

- Users
  - Request data
  - Save data
- Janus Manager
  - Processes requests
  - Interacts with the database
  - Responds with output



Daily activity

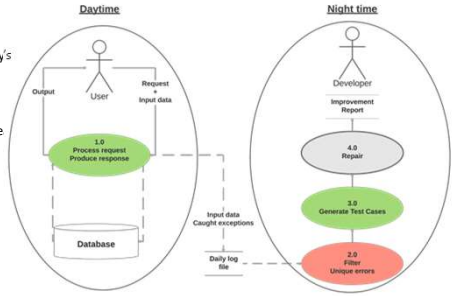
- Procedures in place to catch exceptions and log:
  - Request
  - Input data
  - Type of exception
  - Location of exception
- Logs saved in file on server



Nightly activity

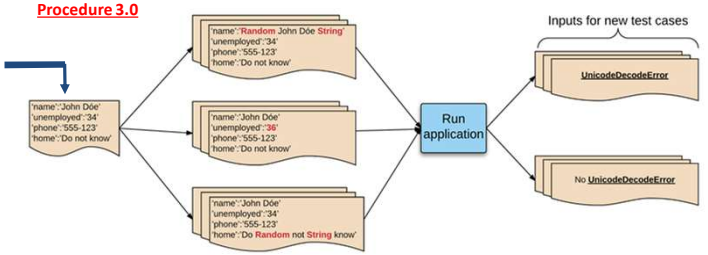
When last user logs out

1. Procedure 2.0 started
   - Sorts and filters the day's exceptions
2. Procedure 3.0
   - Emulates input data, type, size and structure
   - Produces test cases
3. Procedure 4.0
   - Genetic Improvement
   - Parallel process on the server
   - Outputs report for developer



Nightly activity

Procedure 3.0



Improvement log

The story so far

- Janus Manager was developed in March 2016
- Since October 2016 has had GI running as a permanent service
- 22 Bugs reported
  - Variable name mixup - **current_date vs. current_time**
  - Range constants - For i in range(**len(Var)+1**):
  - And more ….
- 22 Bugs fixed
  - Whole process takes 20 minutes (on average)
  - Fix found within 10 generations

Current count is ~40

Daniel's perspective of hyper-heuristics

Real-World Challenges

- Researchers strive to make algorithms increasingly general-purpose
- But practitioners have very specific needs
- Designing custom algorithms tuned to particular problem instance distributions and/or computational architectures can be very time consuming

Automated Design of Algorithms

- Addresses the need for custom algorithms
- But due to high computational complexity, only feasible for repeated problem solving
- Hyper-heuristics accomplish automated design of algorithms by searching program space

Hyper-heuristics

- Hyper-heuristics are a special type of meta-heuristic
  - Step 1: Extract algorithmic primitives from existing algorithms
  - Step 2: Search the space of programs defined by the extracted primitives
- While Genetic Programming (GP) is particularly well suited for executing Step 2, other meta-heuristics can be, and have been, employed
- The type of GP employed matters [24]

Type of GP Matters: Experiment Description

- Implement five types of GP (tree GP, linear GP, canonical Cartesian GP, Stack GP, and Grammatical Evolution) in hyper-heuristics for evolving black-box search algorithms for solving 3-SAT
- Base hyper-heuristic fitness on the fitness of the best search algorithm generated at solving the 3-SAT problem
- Compare relative effectiveness of each GP type as a hyper-heuristic

GP Individual Description

- Search algorithms are represented as an iterative algorithm that passes one or more set of variable assignments to the next iteration
- Genetic program represents a single program iteration
- Algorithm runs starting with a random initial population of solutions for 30 seconds

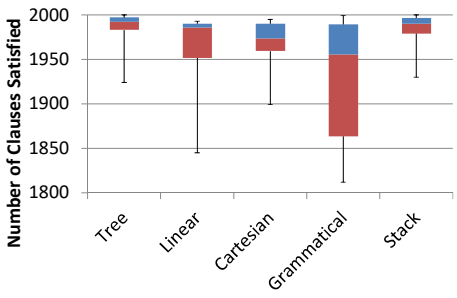
3-SAT Problem

- A subset of the Boolean Satisfiability Problem (SAT)
- The goal is to select values for Boolean variables such that a given Boolean equation evaluates as true (is satisfied)
- Boolean equations are in 3-conjunctive normal form
- Example:
  - (A ∨ B ∨ C) ∧ (¬A ∨ ¬C ∨ D) ∧ (¬B ∨ C ∨ ¬D)
  - Satisfied by ¬A, B, C, ¬D
- Fitness is the number of clauses satisfied by the best solution in the final population

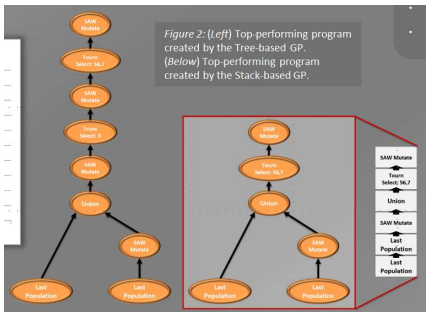
Genetic Programming Nodes Used

- Last population, Random population
- Tournament selection, Fitness proportional selection, Truncation selection, Random selection
- Bitwise mutation, Greedy flip, Quick greedy flip, Stepwise adaption of weights, Novelty
- Union

Results



Results [24]



Results

- Generated algorithms mostly performed comparably well on training and test problems
- Tree and stack GP perform similarly well on this problem, as do linear and Cartesian GP
- Tree and stack GP perform significantly better on this problem than linear and Cartesian GP, which perform significantly better than grammatical evolution

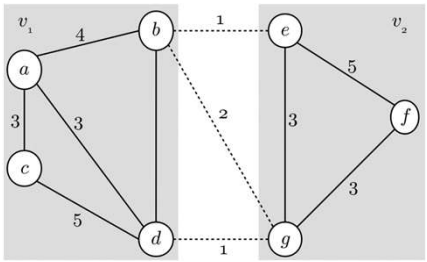
Conclusions

- The choice of GP type makes a significant difference in the performance of the hyper-heuristic
- The size of the search space appears to be a major factor in the performance of the hyper-heuristic

Case Study 1: Evolving Multi-level Graph Partitioning Algorithms [28]

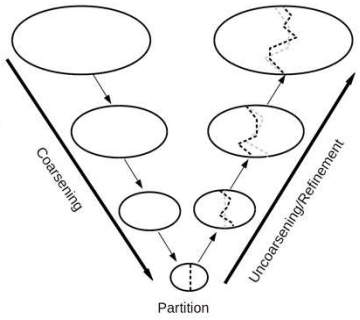
Graph Partitioning

- Graph partitioning can be applied to several problems
  - Parallel computation: minimize inter-process communication
  - Network security: placing network traffic monitors at critical junctures
- In general, optimal graph partitioning is NP-hard
- Real-time applications require fast approximation heuristics

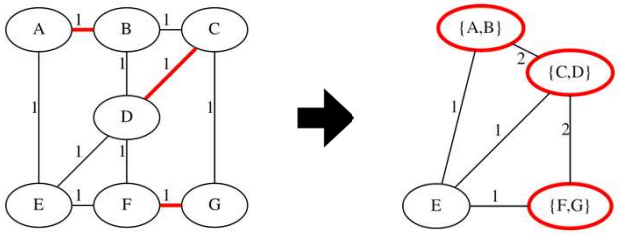


Multi-level Graph Partitioning

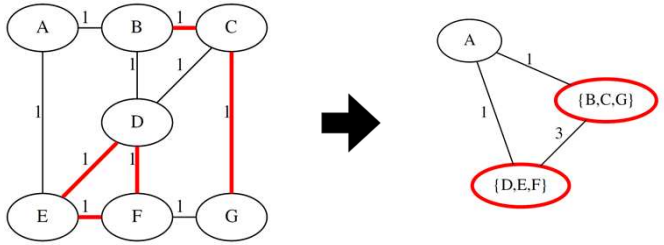
1. Approximate graph with sequence of smaller graphs
2. Quickly partition smallest graph
3. Use partition of smallest graph as an initial partition of the next smallest
4. Perform greedy search to improve the quality of the partition
5. Repeat steps 3 and 4 until partition is found for the original graph



Coarsening: Matching Contraction



Coarsening: Subgraph Contraction



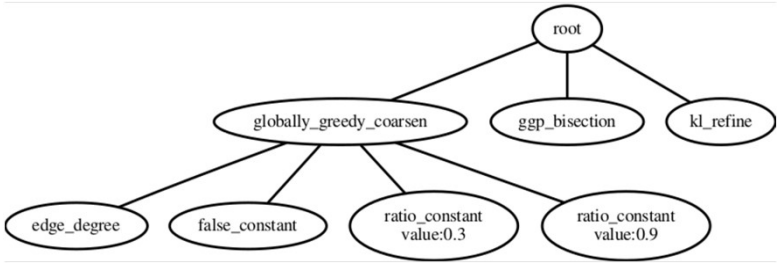
Optimizing Multi-level Partitioning

- Multi-level partitioning of power-law graphs can be improved with specialized coarsening heuristics*
- Demonstrates the potential to customize an algorithm for graphs with particular characteristics
- Manually developing new heuristics can be expensive
- Process can be automated using hyper-heuristics

**A. Abou-Rjeili and G. Karypis, "Multilevel Algorithms for Partitioning Power-law Graphs," in Proceedings of the 20th International Parallel and Distributed Processing Symposium (IPDPS'06), IEEE, Rhodes Island, Greece, 2006.*

Hyper-heuristic Approach

- Extract functionality from existing graph partitioning techniques



Methodology

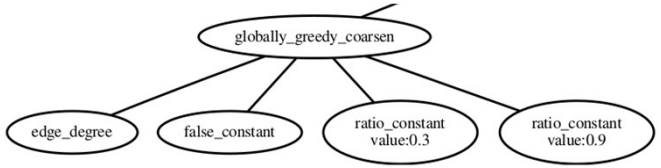
- Evolve a population of graph partition algorithms
- Strongly typed parse tree representation
- Evolution targeted at specific type of graph (e.g., random graph model, computer network)
- Cost of partitions produced determines solution fitness (averaged over many graphs)

$$Fitness = \frac{1}{|P|}\sum_{p\in P}\left[\sum_{(u,v)|p[u]\neq p[v]} w_e[(u,v)]\right]$$

Primitive Operations

- **Root node**
  - child node for each of: coarsening, partitioning, uncoarsening
- **Coarsening**
  - Traditional matching schemes: random, heavy edge, light edge
  - Globally greedy: considers all edges in graph sorted by some metric
  - Locally greedy: randomly visits vertices, considers incident edges
  - Edge metrics: weight, degree, or core number of incident vertices, or edge weight (can be combined using math operators)

Globally and Locally Greedy Coarsening



- Edge metric
- Reverse ordering (maximize)
- Maximum vertex weight ratio
- Maximum contraction ratio

Primitive Operations (cont.)

- **Partitioning**
  - random, growth, greedy growth, spectral, Kernighan-Lin
- **Uncoarsening**
  - KL refinement, greedy refinement
- **Miscellaneous**
  - Constants: integers, ratios, booleans
  - Random conditional: randomly true or false according to child probability value

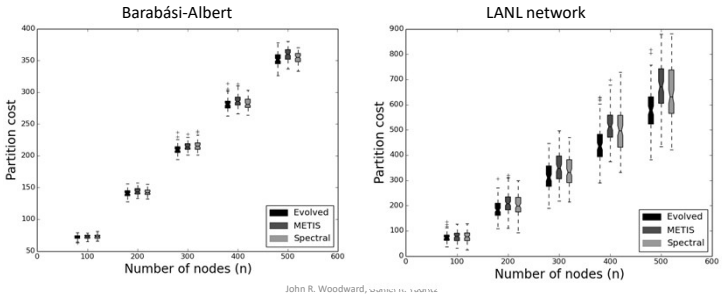
Experiment

- Partition algorithms are evolved targeting:
  - Erdös-Rényi random graphs
  - Barabási-Albert random graphs
  - Los Alamos National Laboratory (LANL) authentication graphs

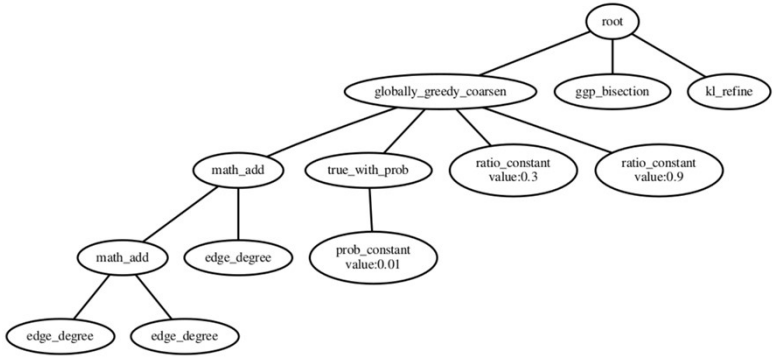
Performance is compared against METIS partitioning software as well as spectral partitioning

Results

Method	E_{ER}	E_{BA}	E_{LANL}	METIS	SP
E_{ER}	0.0	-0.06	-0.38	-0.06	-4.55
E_{BA}	-0.53	0.0	-0.12	-0.72	-0.92
E_{LANL}	-0.48	-0.10	0.0	-2.97	-3.90



Representative High Fitness Final Evolved Solution (E_{BA})



Conclusion

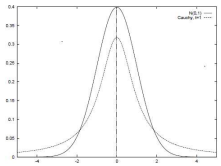
- Multi-level partitioning algorithms can be tailored to specific graph types, including real-world networks
- These customized heuristics improve upon general purpose graph partitioning methods
- The design process can be automated using hyper-heuristics

Case Study 2: The Automated Design of Mutation Operators for Evolutionary Programming

Designing Mutation Operators for Evolutionary Programming [18]

1. **Evolutionary programing** optimizes functions by evolving a population of real-valued vectors (genotype).
2. **Variation** has been provided (manually) by **probability distributions** (**Gaussian, Cauchy, Levy**).
3. We are **automatically generating** probability distributions (using genetic programming).
4. **Not from scratch**, but from already well known distributions (**Gaussian, Cauchy, Levy**). We are "**genetically improving probability distributions**".
5. We are evolving mutation operators **for a problem class** (probability distributions over functions).
6. NO CROSSOVER

Genotype is (1.3,...,4.5,...,8.7) Before mutation



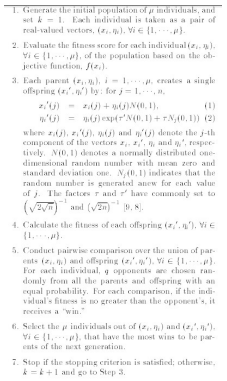
Genotype is (1.2,...,4.4,...,8.6) After mutation

(Fast) Evolutionary Programming

Heart of algorithm is mutation
SO LETS AUTOMATICALLY DESIGN

$$x_i'(j) = x_i(j) + \eta_i(j) D_j$$

1. **EP** mutates with a **Gaussian**
2. **FEP** mutates with a **Cauchy**
3. A **generalization** is mutate with a **distribution D** (generated with genetic programming)



1. Generate the initial population of μ individuals, and set $k = 1$. Each individual is taken as a pair of real-valued vectors, (x_i, η_i), $\forall i \in \{1, \cdots, \mu\}$.
2. Evaluate the fitness score for each individual (x_i, η_i), $\forall i \in \{1, \cdots, \mu\}$, of the population based on the objective function, $f(x_i)$.
3. Each parent (x_i, η_i), $i = 1, \cdots, \mu$, creates a single offspring (x_i', η_i') by: for $j = 1, \cdots, n$,
$$x_i'(j) = x_i(j) + \eta_i(j) N(0,1), \quad (1)$$
$$\eta_i'(j) = \eta_i(j) \exp(\tau' N(0,1) + \tau N_j(0,1)) \quad (2)$$
where $x_i(j)$, $x_i'(j)$, $\eta_i(j)$ and $\eta_i'(j)$ denote the j-th component of the vectors x_i, x_i', η_i and η_i', respectively. $N(0,1)$ denotes a normally distributed one-dimensional random number with mean zero and standard deviation one. $N_j(0,1)$ indicates that the random number is generated anew for each value of j. The factors τ and τ' have commonly set to $\left(\sqrt{2\sqrt{n}}\right)^{-1}$ and $\left(\sqrt{2n}\right)^{-1}$ [9, 8].
4. Calculate the fitness of each offspring (x_i', η_i'), $\forall i \in \{1, \cdots, \mu\}$.
5. Conduct pairwise comparison over the union of parents (x_i, η_i) and offspring (x_i', η_i'), $\forall i \in \{1, \cdots, \mu\}$. For each individual, q opponents are chosen randomly from all the parents and offspring with an equal probability. For each comparison, if the individual's fitness is no greater than the opponent's, it receives a "win."
6. Select the μ individuals out of (x_i, η_i) and (x_i', η_i'), $\forall i \in \{1, \cdots, \mu\}$, that have the most wins to be parents of the next generation.
7. Stop if the stopping criterion is satisfied; otherwise, $k = k + 1$ and go to Step 3.

Optimization & Benchmark Functions

A set of 23 benchmark functions is typically used in the literature. **Minimization** $\forall x \in S : f(x_{min}) \leq f(x)$

We use them as **problem classes**.

Table 1: The 23 test functions used in our experimental studies, where n is the dimension of the function, f_{min} the minimum value of the function, and $S \subseteq R^n$

Test function	n	S	f_{min}
$f_1(x) = \sum_{i=1}^{n} x_i^2$	30	$[-100, 100]^n$	0
$f_2(x) = \sum_{i=1}^{n} \lvert x_i \rvert + \prod_{i=1}^{n} \lvert x_i \rvert$	30	$[-10, 10]^n$	0
$f_3(x) = \sum_{i=1}^{n} (\sum_{j=1}^{i} x_j)^2$	30	$[-100, 100]^n$	0
$f_4(x) = \max_i \{ \lvert x_i \rvert, 1 \leq i \leq n \}$	30	$[-100, 100]^n$	0
$f_5(x) = \sum_{i=1}^{n-1} [100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2]$	30	$[-30, 30]^n$	0
$f_6(x) = \sum_{i=1}^{n} \lfloor x_i + 0.5 \rfloor$	30	$[-100, 100]^n$	0
$f_7(x) = \sum_{i=1}^{n} i x_i^4 + random[0, 1)$	30	$[-1.28, 1.28]^n$	0
$f_8(x) = \sum_{i=1}^{n} -x_i \sin(\sqrt{\lvert x_i \rvert})$	30	$[-500, 500]^n$	-12569.5
$f_9(x) = \sum_{i=1}^{n} [x_i^2 - 10\cos(2\pi x_i) + 10)]$	30	$[-5.12, 5.12]^n$	0
$f_{10}(x) = -20 \exp\left(-0.2\sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}\right) - \exp\left(\frac{1}{n}\sum_{i=1}^{n} \cos 2\pi x_i\right) + 20 + e$	30	$[-32, 32]^n$	0

Function Class 1

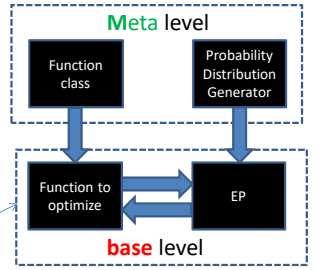
1. Machine learning needs to generalize.
2. We generalize to function classes.
3. y = x^2 (**a function**)
4. y = ax^2(parameterised function)
5. y = ax^2, a ~[1,2] (**function class**)
6. We do this for all benchmark functions.
7. **The mutation operators is evolved to fit the probability distribution of functions**.

Function Classes 2

Function Classes	S	b	f_{min}
$f_1(x) = a\sum_{i=1}^{n} x_i^2$	$[-100, 100]^n$	N/A	0
$f_2(x) = a\sum_{i=1}^{n} \lvert x_i \rvert + b\prod_{i=1}^{n} \lvert x_i \rvert$	$[-10, 10]^n$	$b \in [0, 10^{-5}]$	0
$f_3(x) = \sum_{i=1}^{n} (a\sum_{j=1}^{i} x_j)^2$	$[-100, 100]^n$	N/A	0
$f_4(x) = \max_i \{ a \mid x_i \mid, 1 \leq i \leq n \}$	$[-100, 100]^n$	N/A	0
$f_5(x) = \sum_{i=1}^{n} [a(x_{i+1} - x_i^2)^2 + (x_i - 1)^2]$	$[-30, 30]^n$	N/A	0
$f_6(x) = \sum_{i=1}^{n} (\lfloor ax_i + 0.5 \rfloor)^2$	$[-100, 100]^n$	N/A	0
$f_7(x) = a\sum_{i=1}^{n} ix_i^4 + random[0, 1)$	$[-1.28, 1.28]^n$	N/A	0
$f_8(x) = \sum_{i=1}^{n} -(x_i \sin(\sqrt{\lvert x_i \rvert}) + a)$	$[-500, 500]^n$	N/A	[-12629.5, -12599.5]
$f_9(x) = \sum_{i=1}^{n} [ax_i^2 + b(1 - cos(2\pi x_i))]$	$[-5.12, 5.12]^n$	$b \in [5, 10]$	0
$f_{10}(x) = -a\exp(-0.2\sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2})$ $- \exp(\frac{1}{n}\sum_{i=1}^{n} \cos 2\pi x_i) + a + e$	$[-32, 32]^n$	N/A	0

Meta and Base Learning

- At the **base** level we are learning about a **specific** function.
- At the **meta** level we are learning about the problem **class**.
- We are just doing **"generate and test"** at a higher level
- What is being passed with each **blue arrow**?
- **Conventional** EP



Compare Signatures (Input-Output)

Evolutionary Programming	Evolutionary Programming Designer
$(R^n \rightarrow R) \rightarrow R^n$	$[(R^n \rightarrow R)] \rightarrow ((R^n \rightarrow R) \rightarrow R^n)$
Input is a function mapping real-valued vectors of length n to a real-value.	**Input** is a *list of* functions mapping real-valued vectors of length n to a real-value (i.e. sample problem instances from the problem class).
Output is a (near optimal) real-valued vector	**Output** is a (near optimal) (mutation operator for) Evolutionary Programming
(i.e. the solution to the problem instance)	(i.e. the solution method to the problem class)

We are **raising the level of generality** at which we operate.

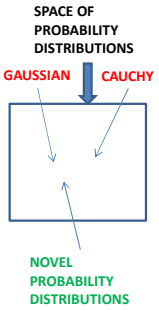
Genetic Programming to Generate Probability Distributions

1. GP **Function Set** {+, -, *, %}
2. GP **Terminal Set** {N(0, random)}

N(0,1) is a normal distribution.

For example a Cauchy distribution is generated by N(0,1)%N(0,1).

Hence **the search space of probability distributions** contains the two existing probability distributions used in EP but also **novel probability distributions**.



Means and Standard Deviations

These results are good for two reasons.
1. starting with a manually designed distributions (Gaussian).
2. evolving distributions **for each function class**.

Function Class	FEP Mean Best	FEP Std Dev	CEP Mean Best	CEP Std Dev	GP-distribution Mean Best	GP-distribution Std Dev
f_1	1.24×10^{-3}	2.69×10^{-4}	1.45×10^{-4}	9.95×10^{-5}	$\mathbf{6.37\times10^{-5}}$	5.56×10^{-5}
f_2	1.53×10^{-1}	2.72×10^{-2}	4.30×10^{-2}	9.08×10^{-3}	$\mathbf{8.14\times10^{-4}}$	8.50×10^{-4}
f_3	2.74×10^{-2}	2.43×10^{-2}	5.15×10^{-2}	9.52×10^{-2}	$\mathbf{6.14\times10^{-3}}$	8.78×10^{-3}
f_4	1.79	1.84	1.75×10	6.10	$\mathbf{2.16\times10^{-1}}$	6.54×10^{-1}
f_5	2.52×10^{-3}	4.96×10^{-4}	2.66×10^{-4}	4.65×10^{-5}	$\mathbf{8.39\times10^{-7}}$	1.43×10^{-7}
f_6	3.86×10^{-2}	3.12×10^{-2}	4.40×10	1.42×10^{2}	$\mathbf{9.20\times10^{-3}}$	1.34×10^{-2}
f_7	6.49×10^{-2}	1.04×10^{-2}	6.64×10^{-2}	1.21×10^{-2}	$\mathbf{5.25\times10^{-2}}$	8.46×10^{-3}
f_8	-11342.0	3.26×10^{2}	-7894.6	6.14×10^{2}	**-12611.6**	2.30×10
f_9	6.24×10^{-2}	1.30×10^{-2}	1.09×10^{2}	3.58×10	$\mathbf{1.74\times10^{-3}}$	4.25×10^{-4}
f_{10}	1.67	4.26×10^{-1}	1.45	2.77×10^{-1}	**1.38**	2.45×10^{-1}

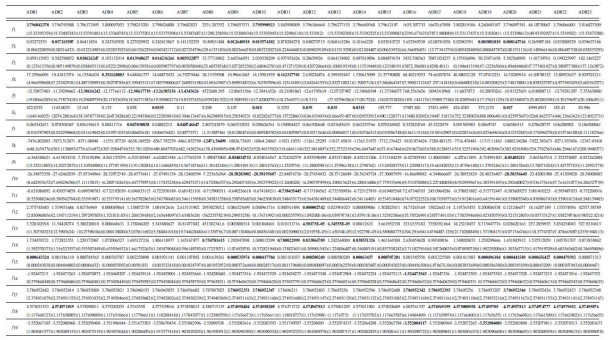
T-tests

Table 5 2-tailed t-tests comparing EP with GP-distributions, FEP and CEP on f_1-f_{10}.

Function Class	Number of Generations	GP-distribution vs FEP *t-test*	GP-distribution vs CEP *t-test*
f_1	1500	2.78×10^{-47}	4.07×10^{-2}
f_2	2000	5.53×10^{-62}	1.59×10^{-54}
f_3	5000	8.03×10^{-8}	1.14×10^{-3}
f_4	5000	1.28×10^{-7}	3.73×10^{-36}
f_5	20000	2.80×10^{-58}	9.29×10^{-63}
f_6	1500	1.85×10^{-8}	3.11×10^{-2}
f_7	3000	3.27×10^{-9}	2.00×10^{-9}
f_8	9000	7.99×10^{-48}	5.82×10^{-75}
f_9	5000	6.37×10^{-55}	6.54×10^{-39}
f_{10}	1500	9.23×10^{-5}	1.93×10^{-1}

Performance on Other Problem Classes

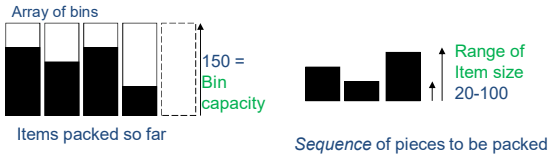
Table 8: This table compares the fitness values (averaged over 20 runs) of each of the 23 ADRs on each of the 23 function classes. Standard deviations are in parentheses.



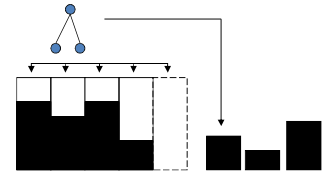
Case Study 3: The Automated Design of On-Line Bin Packing Algorithms

On-line Bin Packing Problem [9,11]

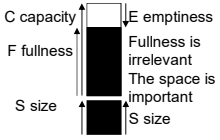
- A *sequence* of items packed into as few a bins as possible.
- Bin size is 150 units, items uniformly distributed between 20-100.
- Different to the off-line bin packing problem where the *set* of items.
- The "best fit" heuristic, places the current item in the space it fits best (leaving least slack).
- It has the property that this heuristic does not open a new bin unless it is forced to.



Genetic Programming applied to on-line bin packing

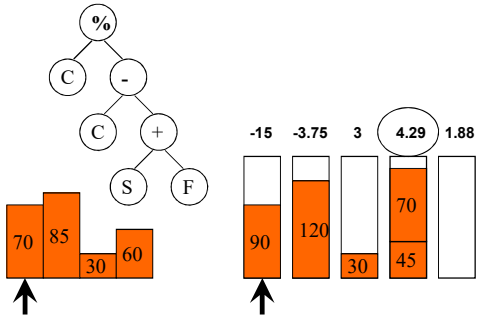


Not obvious how to link Genetic Programming to combinatorial problems.
The GP tree is applied to each bin with the current item and placed in the bin with
The maximum score



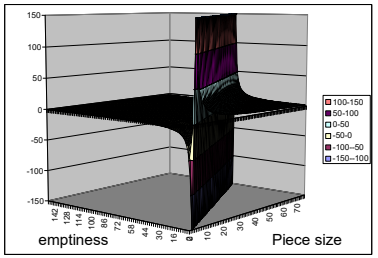
Terminals supplied to Genetic Programming
Initial representation {C, F, S}
Replaced with {E, S}, E=C-F

How the heuristics are applied (skip)

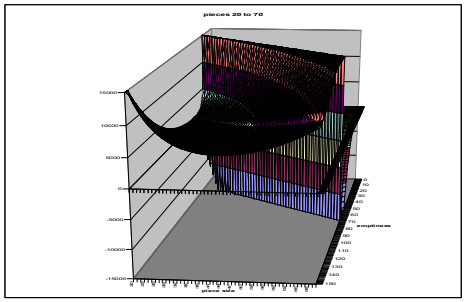


The Best Fit Heuristic

Best fit = 1/(E-S). Point out features.
Pieces of size S, which fit well into the space remaining E, score well.
Best fit applied produces a set of points on the surface,
The bin corresponding to the maximum score is picked.



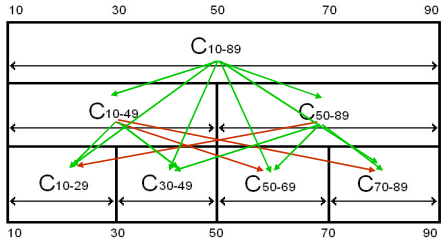
Our Best Heuristic



Similar shape to best fit – but curls up in one corner.
Note that this is rotated, relative to previous slide.

Robustness of Heuristics

= all legal results
= some illegal results



Testing Heuristics on problems of much larger size than in training

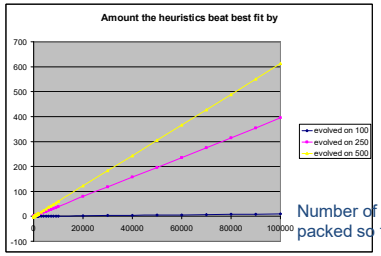
Table I	*H trained100*	*H trained 250*	*H trained 500*
100	0.427768358	0.298749035	0.140986023
1000	0.406790534	0.010006408	0.000350265
10000	0.454063071	2.58E-07	9.65E-12
100000	0.271828318	1.38E-25	2.78E-32

Table shows p-values using the best fit heuristic, for heuristics trained on different size problems, when applied to different sized problems

1. As number of items trained on increases, the probability decreases (see next slide).
2. As the number of items packed increases, the probability decreases (see next slide).

Compared with Best Fit

Amount evolved heuristics beat best fit by.

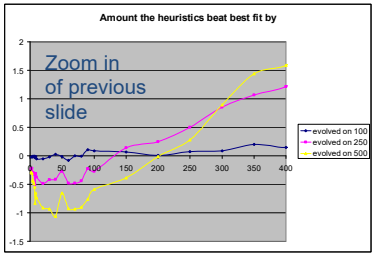


Number of pieces packed so far.

- Averaged over 30 heuristics over 20 problem instances
- Performance does not deteriorate
- The larger the training problem size, the better the bins are packed.

Compared with Best Fit

Amount evolved heuristics beat best fit by.



- The heuristic seems to learn the number of pieces in the problem
- Analogy with sprinters running a race – accelerate towards end of race.
- The "break even point" is approximately half of the size of the training problem size
- If there is a gap of size 30 and a piece of size 20, it would be better to wait for a better piece to come along later – about 10 items (similar effect at upper bound?).

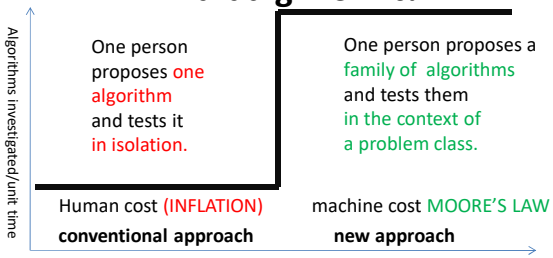
Step by Step Guide to Automatic Design of Algorithms [8, 12]

1. Study the literature for existing heuristics for your chosen domain (manually designed heuristics).
2. Build an algorithmic framework or template which expresses the known heuristics.
3. Let metaheuristics (e.g., Genetic Programming) search for *variations on the theme*.
4. Train and test on problem instances drawn from the same probability distribution (like machine learning). Constructing an optimizer is machine learning (this approach prevents "cheating").

A Brief History (Example Applications) [5]

1. **Image Recognition** – Roberts Mark
2. **Travelling Salesman Problem** – Keller Robert
3. **Boolean Satisfiability** – Holger Hoos, Fukunaga, Bader-El-Den, Alex Bertels & Daniel Tauritz
4. **Data Mining** – Gisele L. Pappa, Alex A. Freitas
5. **Decision Tree** - Gisele L. Pappa et al
6. **Crossover Operators** – Oltean et al, Brian Goldman and Daniel Tauritz
7. **Selection Heuristics** – Woodward & Swan, Matthew Martin & Daniel Tauritz
8. **Bin Packing 1,2,3 dimension** (on and off line) Edmund Burke et. al. & Riccardo Poli et al
9. **Bug Location –** Shin Yoo
10. **Job Shop Scheduling –** Mengjie Zhang
11. **Black Box Search Algorithms** – Daniel Tauritz et al
12. **Graph Algorithms** – Aaron Pope & Daniel Tauritz

A Paradigm Shift?



- Previously **one** person proposes **one** algorithm
- Now **one** person proposes **a set of** algorithms
- Analogous to "industrial revolution" from hand made to machine made. Automatic Design.

Conclusions

1. Heuristic are **trained to fit a problem class**, so are designed in context (like evolution). Let's close the feedback loop! Problem instances live in classes.
2. We can design algorithms on **small** problem instances and **scale** them apply them to **large** problem instances (TSP, child multiplication).

SUMMARY

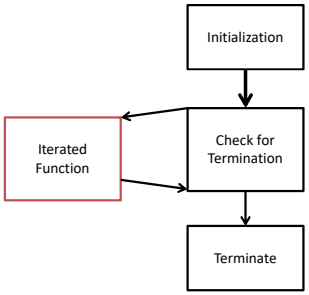
1. We can automatically design algorithms that **consistently outperform human designed algorithms (on various domains)**.
2. The "best" heuristics depends on the set of problem instances. (feedback)
3. Resulting algorithm is part man-made part machine-made (synergy)
4. not evolving from scratch like Genetic Programming,
5. improve existing algorithms and adapt them to the new problem instances.
6. Algorithms are **reusable**, "solutions" aren't. (e.g. tsp algorithm vs route)

Case Study 4: The Automated Design of Black Box Search Algorithms [21, 23, 25]

Approach

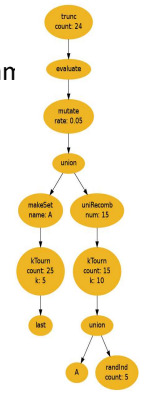
- Hyper-Heuristic employing Genetic Programing
- Post-ordered parse tree
- Evolve the iterated function

Our Solution



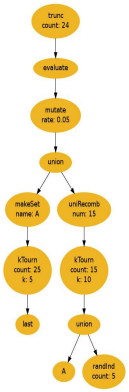
Our Solution

- Hyper-Heuristic employing Genetic Program
- Post-ordered parse tree
- Evolve the iterated function
- High-level primitives



Parse Tree

- Iterated function
- Sets of solutions
- Function returns a set of solutions accessible to the next iteration

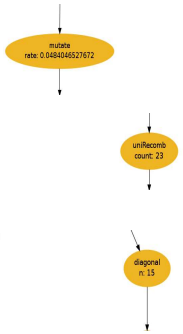


Primitive Types

- Variation Primitives
- Selection Primitives
- Set Primitives
- Evaluation Primitive
- Terminal Primitives

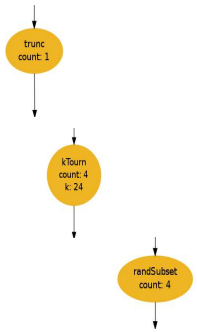
Variation Primitives

- Bit-flip Mutation
  - *rate*
- Uniform Recombination
  - *count*
- Diagonal Recombination
  - *n*

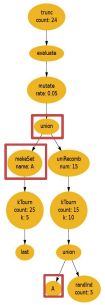


Selection Primitives

- Truncation Selection
  - *count*
- K-Tournament Selection
  - *k*
  - *count*
- Random Sub-set Selection
  - *count*



Set-Operation Primitives



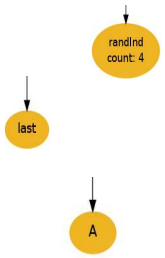
- Make Set
  - *name*
- Persistent Sets
  - *name*
- Union

Evaluation Primitive

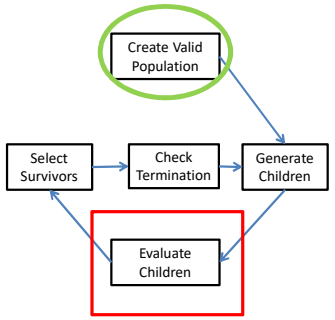
- Evaluates the nodes passed in
- Allows multiple operations and accurate selections within an iteration
  - Allows for deception

Terminal Primitives

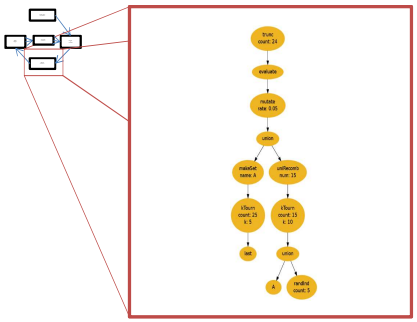
- Random Individuals
  - *count*
- `Last' Set
- Persistent Sets
  - *name*



Meta-Genetic Program



BBSA Evaluation

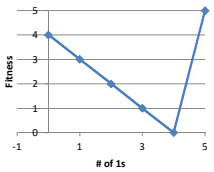


Termination Conditions

- Evaluations
- Iterations
- Operations
- Convergence

Proof of Concept Testing

- Deceptive Trap Problem



Proof of Concept Testing (cont.)

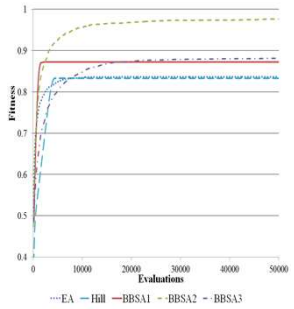
- Evolved Problem Configuration
  - Bit-length = 100
  - Trap Size = 5
- Verification Problem Configurations
  - Bit-length = 100, Trap Size = 5
  - Bit-length = 200, Trap Size = 5
  - Bit-length = 105, Trap Size = 7
  - Bit-length = 210, Trap Size = 7

Results

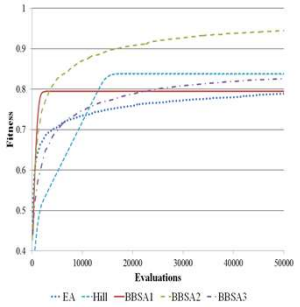
BBSA	EA	Hill-Climber
1	+	+
2	+	+
3	+	+
4	-	-
5	+	+
6	+	+
7	+	+
8	-	-
9	-	-
10	-	-
11	+	+
12	-	-
13	+	+
14	+	+
15	-	-

60% Success Rate

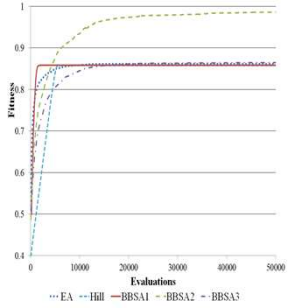
Results:
Bit-Length = 100
Trap Size = 5



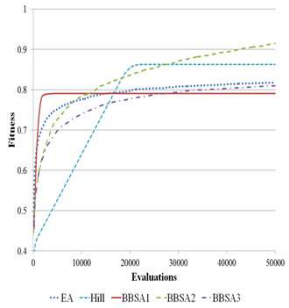
Results:
Bit-Length = 200
Trap Size = 5



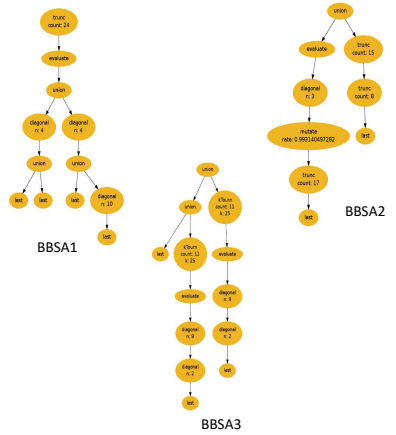
Results:
Bit-Length = 105
Trap Size = 7



Results:
Bit-Length = 210
Trap Size = 7

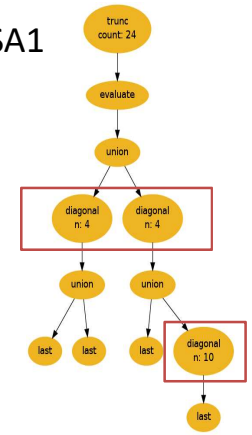


John R. Woodward, Daniel R. Tauritz



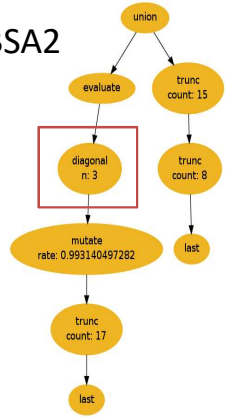
John R. Woodward, Daniel R. Tauritz

BBSA1



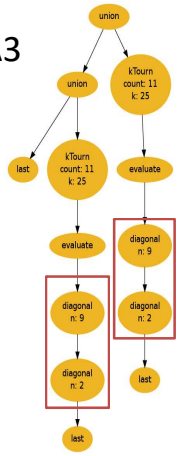
John R. Woodward, Daniel R. Tauritz

BBSA2



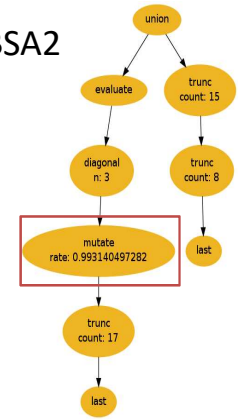
John R. Woodward, Daniel R. Tauritz

BBSA3



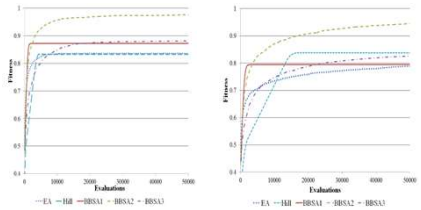
John R. Woodward, Daniel R. Tauritz

BBSA2



John R. Woodward, Daniel R. Tauritz

Over-Specialization

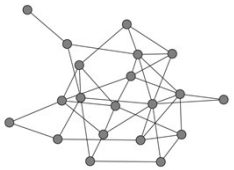


Trained Problem Configuration

Alternate Problem Configuration

John R. Woodward, Daniel R. Tauritz

Case Study 5: Evolving Random Graph Generators: A Case for Increased Algorithmic Primitive Granularity [27]



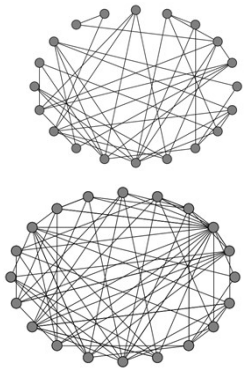
John R. Woodward, Daniel R. Tauritz

Random Graphs

- Graphs are a powerful modeling tool
  - Computer and social networks
  - Transportation and power grids
- Algorithms designed for graphs
  - Community detection and graph partitioning
  - Network routing and intrusion detection
- Random graphs provide test data
- Prediction using random graphs
  - Spread of disease
  - Deployment of wireless sensors

Traditional Random Graph Models

- Erdös-Rényi
- Barabási-Albert



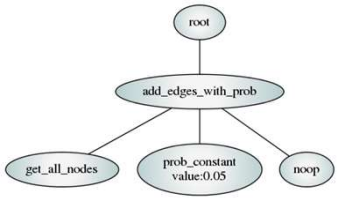
Automated Random Graph Model Design

- Random graph model needs to accurately reflect intended concept
- Model selection can be automated, but relies on having a good solution available
- Developing an accurate model for a new application can be difficult

Can the model design process be automated to produce an accurate graph model given examples?

Hyper-heuristic Approach

- Extract functionality from existing graph generation techniques
- Use Genetic Programming (GP) to construct new random graph algorithms



Previous Attempts at Evolving Random Graph Generators

- Assumes “growth” model, adding one node at a time
- Does well at reproducing traditional models
- Not demonstrated to do well at generating real complex networks
- Limits the search space of possible solutions

Increased Algorithmic Primitive Granularity

- Remove the assumed “growth” structure
- More flexible lower-level primitive set
- Benefit: Can represent a larger variety of algorithms
- Drawback: Larger search space, increasing complexity

Methodology

- NSGA-II evolves population of random graph models
- Strongly typed parse tree representation
- Centrality distributions used to evaluate solution
- quality (degree, betweenness, PageRank)

Primitive Operations

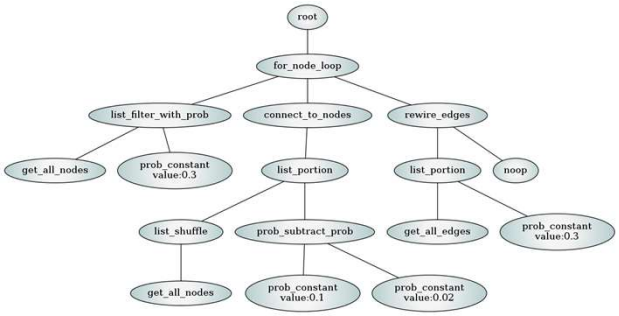
Terminals

- Graph elements: nodes, edges
- Graph properties: average degree, size, order
- Constants: integers, probabilities, Booleans, user inputs
- No-op terminators

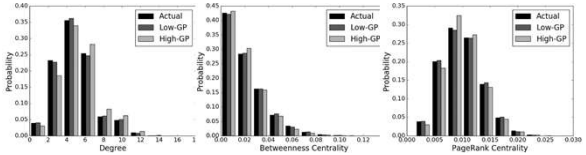
Functions

- Basic programming constructs: for, while, if-else
- Data structures: lists of values, nodes, or edges, list
- combining/selection/sorting
- Math and logic operators: add, multiply, <, ==, AND, OR
- Graph operators: add edges, add subgraph, rewire edges

Example Evolved Random Graph Generator



Reproducing Erdös-Rényi

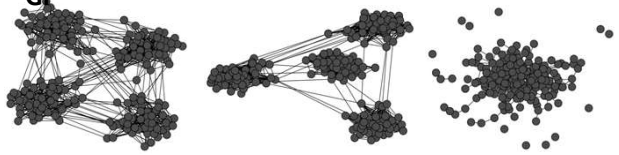


	Low-GP			High-GP	
Metric	**Mean**	**σ**	**Comparison**	**Mean**	**σ**
Degree	0.101	0.048	=	0.108	0.047
Betweenness	0.104	0.031	=	0.105	0.033
PageRank	0.110	0.032	=	0.112	0.029

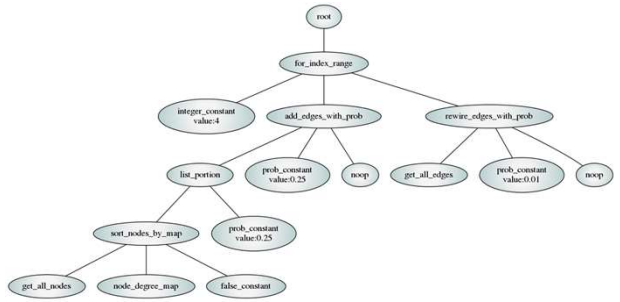
Reproducing Random Community Graphs

	Low-GP			High-GP	
Metric	**Mean**	**σ**	**Comparison**	**Mean**	**σ**
Degree	0.436	0.075	<	0.458	0.055
Betweenness	0.209	0.105	<	0.320	0.126
PageRank	0.127	0.029	<	0.150	0.036

Actual Graph **Low-GP** **High-GP**



Evolved Random Collaboration Network Generator



Conclusion

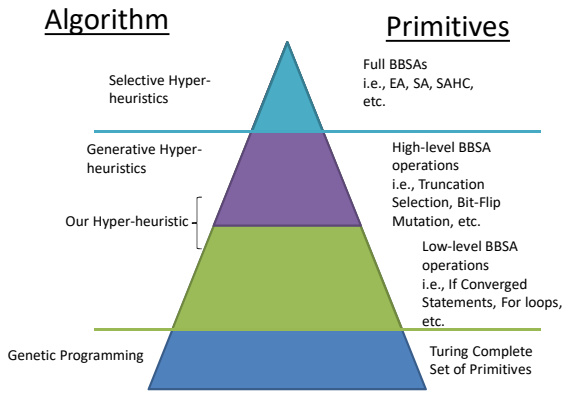
- Traditional random graph models do not always produce appropriate representations of certain concepts
- Accurate random graph model design can be automated using genetic programming
- More flexible set of low-level primitive operations increases resulting model accuracy
- Increase in a priori evolution time is amortized over repeated use of the evolved solutions

Some Final Thoughts

Challenges in Hyper-heuristics

- Hyper-heuristics are very computationally expensive (use Asynchronous Parallel GP [26,30])
- What is the best primitive granularity? (see next slide; also see [41])
- How to automate decomposition and recomposition of primitives?
- How to automate primitive extraction?
- How does hyper-heuristic performance scale for increasing primitive space size? (see [25,27])

Primitive Granularity



End of File ☺

- Thank you for listening !!!
- We are glad to take any
  - – comments (+,-)
  - – suggestions/criticisms

  Please email us any missing references!

  John Woodward (https://www.eecs.qmul.ac.uk/~jwoodward/)

  Daniel Tauritz (https://bonsai.auburn.edu/dtauritz/)

References 1

1. John Woodward. Computable and Incomputable Search Algorithms and Functions. IEEE International Conference on Intelligent Computing and Intelligent Systems (IEEE ICIS 2009), pp. 871-875, 2009.
2. John Woodward. The Necessity of Meta Bias in Search Algorithms. International Conference on Computational Intelligence and Software Engineering (CiSE), pp. 1-4, 2010.
3. John Woodward & Ruibin Bai. Why Evolution is not a Good Paradigm for Program Induction: A Critique of Genetic Programming. In Proceedings of the first ACM/SIGEVO Summit on Genetic and Evolutionary Computation, pp. 593-600, 2009.
4. Jerry Swan, John Woodward, Ender Ozcan, Graham Kendall, Edmund Burke. Searching the Hyper-heuristic Design Space. Cognitive Computation, 6:66-73, 2014.
5. Gisele L. Pappa, Gabriela Ochoa, Matthew R. Hyde, Alex A. Freitas, John Woodward, Jerry Swan. Contrasting meta-learning and hyper-heuristic research. Genetic Programming and Evolvable Machines, 15:3-35, 2014.
6. Edmund K. Burke, Matthew Hyde, Graham Kendall, and John Woodward. Automating the Packing Heuristic Design Process with Genetic Programming. Evolutionary Computation, 20(1):63-89, 2012.
7. Edmund K. Burke, Matthew R. Hyde, Graham Kendall, and John Woodward. A Genetic Programming Hyper-Heuristic Approach for Evolving Two Dimensional Strip Packing Heuristics. IEEE Transactions on Evolutionary Computation, 14(6):942-958, 2010.
8. Edmund K. Burke, Matthew R. Hyde, Graham Kendall, Gabriela Ochoa, Ender Ozcan and John R. Woodward. Exploring Hyper-heuristic Methodologies with Genetic Programming, Computational Intelligence: Collaboration, Fusion and Emergence, In C. Mumford and L. Jain (eds.), Intelligent Systems Reference Library, Springer, pp. 177-201, 2009.
9. Edmund K. Burke, Matthew Hyde, Graham Kendall and John R. Woodward. The Scalability of Evolved On Line Bin Packing Heuristics. In Proceedings of the IEEE Congress on Evolutionary Computation, pp. 2530-2537, 2007.
10. R. Poli, John R. Woodward, and Edmund K. Burke. A Histogram-matching Approach to the Evolution of Bin-packing Strategies. In Proceedings of the IEEE Congress on Evolutionary Computation, pp. 3500-3507, 2007.
11. Edmund K. Burke, Matthew Hyde, Graham Kendall, and John Woodward. Automatic Heuristic Generation with Genetic Programming: Evolving a Jack-of-all-Trades or a Master of One, In Proceedings of the Genetic and Evolutionary Computation Conference, pp. 1559-1565, 2007.

References 2

12. John R. Woodward and Jerry Swan. Template Method Hyper-heuristics, Metaheuristic Design Patterns (MetaDeeP) workshop, GECCO Comp'14, pp. 1437-1438, 2014.
13. Saemundur O. Haraldsson and John R. Woodward, Automated Design of Algorithms and Genetic Improvement: Contrast and Commonalities, 4th Workshop on Automatic Design of Algorithms (ECADA), GECCO Comp '14, pp. 1373-1380, 2014.
14. John R. Woodward, Simon P. Martin and Jerry Swan. Benchmarks That Matter For Genetic Programming, 4th Workshop on Evolutionary Computation for the Automated Design of Algorithms (ECADA), GECCO Comp '14, pp. 1397-1404, 2014.
15. John R. Woodward and Jerry Swan. The Automatic Generation of Mutation Operators for Genetic Algorithms, 2nd Workshop on Evolutionary Computation for the Automated Design of Algorithms (ECADA), GECCO Comp' 12, pp. 67-74, 2012.
16. John R. Woodward and Jerry Swan. Automatically Designing Selection Heuristics. 1st Workshop on Evolutionary Computation for Designing Generic Algorithms, pp. 583-590, 2011.
17. Edmund K. Burke, Matthew Hyde, Graham Kendall, Gabriela Ochoa, Ender Ozcan, and John Woodward. A Classification of Hyper-heuristics Approaches, Handbook of Metaheuristics, pp. 449-468, International Series in Operations Research & Management Science, M. Gendreau and J-Y Potvin (Eds.), Springer, 2010.
18. Libin Hong and John Woodward and Jingpeng Li and Ender Ozcan. Automated Design of Probability Distributions as Mutation Operators for Evolutionary Programming Using Genetic Programming. Proceedings of the 16th European Conference on Genetic Programming (EuroGP 2013), volume 7831, pp. 85-96, 2013.
19. Ekaterina A. Smorodkina and Daniel R. Tauritz. Toward Automating EA Configuration: the Parent Selection Stage. In Proceedings of CEC 2007 - IEEE Congress on Evolutionary Computation, pp. 63-70, 2007.
20. Brian W. Goldman and Daniel R. Tauritz. Self-Configuring Crossover. In Proceedings of the 13th Annual Conference Companion on Genetic and Evolutionary Computation (GECCO '11), pp. 575-582, 2011.
21. Matthew A. Martin and Daniel R. Tauritz. Evolving Black-Box Search Algorithms Employing Genetic Programming. In Proceedings of the 15th Annual Conference Companion on Genetic and Evolutionary Computation (GECCO '13), pp. 1497-1504, 2013.

References 3

22. Nathaniel R. Kamrath, Brian W. Goldman and Daniel R. Tauritz. Using Supportive Coevolution to Evolve Self-Configuring Crossover. In Proceedings of the 15th Annual Conference Companion on Genetic and Evolutionary Computation (GECCO '13), pp. 1489-1496, 2013.
23. Matthew A. Martin and Daniel R. Tauritz. A Problem Configuration Study of the Robustness of a Black-Box Search Algorithm Hyper-Heuristic. In Proceedings of the 16th Annual Conference Companion on Genetic and Evolutionary Computation (GECCO '14), pp. 1389-1396, 2014.
24. Sean Harris, Travis Bueter, and Daniel R. Tauritz. A Comparison of Genetic Programming Variants for Hyper-Heuristics. In Proceedings of the 17th Annual Conference Companion on Genetic and Evolutionary Computation (GECCO '15), pp. 1043-1050, 2015.
25. Matthew A. Martin and Daniel R. Tauritz. Hyper-Heuristics: A Study On Increasing Primitive-Space. In Proceedings of the 17th Annual Conference Companion on Genetic and Evolutionary Computation (GECCO '15), pp. 1051-1058, 2015.
26. Alex R. Bertels and Daniel R. Tauritz. Why Asynchronous Parallel Evolution is the Future of Hyper-heuristics: A CDCL SAT Solver Case Study. In Proceedings of the 18th Annual Conference Companion on Genetic and Evolutionary Computation (GECCO `16), pp. 1359-1365, 2016.
27. Aaron S. Pope, Daniel R. Tauritz and Alexander D. Kent. Evolving Random Graph Generators: A Case for Increased Algorithmic Primitive Granularity. In Proceedings of the 2016 IEEE Symposium Series on Computational Intelligence (IEEE SSCI 2016), 2016.
28. Aaron S. Pope, Daniel R. Tauritz and Alexander D. Kent. Evolving Multi-level Graph Partitioning Algorithms. In Proceedings of the 2016 IEEE Symposium Series on Computational Intelligence (IEEE SSCI 2016), 2016.
29. Islam Elnabarawy, Daniel R. Tauritz, Donald C. Wunsch. Evolutionary Computation for the Automated Design of Category Functions for Fuzzy ART: An Initial Exploration. In Proceedings of the 19th Annual Conference Companion on Genetic and Evolutionary Computation (GECCO'17), pp. 1133-1140, 2017.
30. Adam Harter, Daniel R. Tauritz, William M. Siever. Asynchronous Parallel Cartesian Genetic Programming. In Proceedings of the 19th Annual Conference Companion on Genetic and Evolutionary Computation (GECCO'17), pp. 1820-1824, 2017.

References 4

31. Marketa Illetskova, Alex R. Bertels, Joshua M. Tuggle, Adam Harter, Samuel Richter, Daniel R. Tauritz, Samuel Mulder, Denis Bueno, Michelle Leger and William M. Siever. Improving Performance of CDCL SAT Solvers by Automated Design of Variable Selection Heuristics. In Proceedings of the 2017 IEEE Symposium Series on Computational Intelligence (SSCI 2017), 2017.
32. John R. Woodward, Jerry Swan, "Why classifying search algorithms is essential", Progress in Informatics and Computing (PIC) 2010 IEEE International Conference on, vol. 1, pp. 285-289, 2010.
33. Samuel N. Richter and Daniel R. Tauritz. The Automated Design of Probabilistic Selection Methods for Evolutionary Algorithms. In Proceedings of the 20th Annual Conference Companion on Genetic and Evolutionary Computation (GECCO 2018), pp. 1545-1552, 2018.
34. Aaron Scott Pope, Robert Morning, Daniel R. Tauritz, and Alexander D. Kent. Automated Design of Network Security Metrics. In Proceedings of the 20th Annual Conference Companion on Genetic and Evolutionary Computation (GECCO 2018), pp. 1680-1687, 2018.
38. John R. Woodward and Ruibin Bai. Canonical representation genetic programming. In Proceedings of the first ACM/SIGEVO Summit on Genetic and Evolutionary Computation, pp. 585-592, 2009.
39. Saemundur O. Haraldsson, John R. Woodward, Alexander EI Brownlee, and Kristin Siggeirsdottir. Fixing bugs in your sleep: How genetic improvement became an overnight success. In Proceedings of the Genetic and Evolutionary Computation Conference Companion, pp. 1513-1520, 2017.
40. Aaron Scott Pope, Daniel R. Tauritz, and Melissa Turcotte. Automated Design of Tailored Link Prediction Heuristics for Applications in Enterprise Network Security. In Proceedings of the 21st Annual Conference Companion on Genetic and Evolutionary Computation (GECCO '19), pp. 1634–1642, 2019.
41. Adam Harter, Aaron Scott Pope, Daniel R. Tauritz, and Chris Rawlings. Empirical Evidence of the Effectiveness of Primitive Granularity Control for Hyper-Heuristics. In Proceedings of the 21st Annual Conference Companion on Genetic and Evolutionary Computation (GECCO '19), pp. 1478–1486, 2019.
42. Aaron Scott Pope, Daniel R. Tauritz, and Chris Rawlings. Automated Design of Random Dynamic Graph Models. In Proceedings of the 21st Annual Conference Companion on Genetic and Evolutionary Computation (GECCO '19), pp. 1504–1512, 2019.
43. Samuel N. Richter, Michael G. Schoen, and Daniel R. Tauritz. Evolving Mean-Update Selection Methods for CMA-ES. In Proceedings of the 21st Annual Conference Companion on Genetic and Evolutionary Computation (GECCO '19), pp. 1513–1517, 2019.

References 5

44. Aaron Scott Pope and Daniel R. Tauritz. Automated Design of Multi-Level Network Partitioning Heuristics Employing Self-Adaptive Primitive Granularity Control. In Proceedings of the 22nd Annual Conference on Genetic and Evolutionary Computation (GECCO '20), 1168–1176, Cancún, Mexico, July 8-12, 2020.
45. Braden N. Tisdale, Aaron Scott Pope, and Daniel R. Tauritz. Dynamic Primitive Granularity Control: An Exploration of Unique Design Considerations. In Proceedings of the 22nd Annual Conference Companion on Genetic and Evolutionary Computation (GECCO '20), pages 1906– 1914, Cancún, Mexico, July 8-12, 2020.
46. Nathaniel R. Kamrath, Aaron Scott Pope, and Daniel R. Tauritz. The Automated Design of Local Optimizers for Memetic Algorithms Employing Supportive Coevolution. In Proceedings of the 22nd Annual Conference Companion on Genetic and Evolutionary Computation (GECCO '20), pages 1889–1897, Cancún, Mexico, July 8-12, 2020.