Using Evolutionary Algorithms
To Solve Hard Problems

(ralph.morelli@trincoll.edu)
• What are Genetic Algorithms.

• Some examples from different domains.

• Implications for secularism? (discussion)

• Complexity: emergent property.

• Simple model: Random variation plus selection.

• GA: inspired by natural selection.

Themes

• Watches without watch makers.

• Implications for secularism? (discussion)
What are GAs?
GAs Competitive with Humans

- AI researchers keep score (Turing Test)

Some examples from Genetic Programming (out of 36):
- Synthesis of a NAND circuit.
- Rediscovery of Cauer elliptical topology for filters.
- Creation of quantum algorithm for Grover’s DB search.
- Creation of quantum algorithm for ‘early promise’ problem.
- Creation of quantum algorithm for ‘early promise’ problem.


All researchers keep score (Turing Test)
Solving Short Cryptograms

• Short cryptograms are hard.
  - Theoretical minimum length ~ 28 characters
  - Unsolvable if too short
  - Possible keys: 26! = 4.0329 × 10^{26}
  - ZWUSQOMKIGECABDFHJLNPRTVXY
  - Key: A permutation of the alphabet, a..z.
  - Simple substitution cipher: a/e, b/g, c/t, ..., x/c, y/m, z/j

• To we or not to we that is the question

Xabcx = River = Raver = Rover = Saves = David...

So long Short Cryptograms
Searching a Problem Space

- **Algorithmic search**
  - Try every key (brute force).
  - Guaranteed to work (theoretically).
  - Intractable: \( \frac{10^{26}}{10^9} \approx 10^9 \) years

- **Heuristic search**
  - Use a rule of thumb to reduce search space.
  - Not guaranteed to succeed.
  - Not guaranteed to work (theoretically).
  - Try every key (brute force).

- **Genetic algorithm**
  - E.G., Traveling salesman nearest neighbor heuristic.

- **Hill climbing optimization**
  - E.G., Hill climbing nearest neighbor heuristic.

- **Genetic algorithm**
  - E.G., Hill climbing optimization.

- **Nearest neighbor heuristic**
  - E.G., Traveling salesman nearest neighbor heuristic.

---

Seaching a Problem Space
Hill Climbing vs GA

Hill Climber
Search Space
Fitness

Genetic Algorithm
Search Space
Fitness

Hill Climbing vs GA
The Genetic Algorithm

Randomly generate an initial Population $P_0$.

Crossover: mate random individuals

Select the fittest individuals

Introduce random mutations

Calculate fitness of $P_t$

$t+1$

Terminating?

Yes

No
GA Basics

Population of Chromosomes

Fitness:

Crossover

Mutation
Crossover and Mutation
### Results

<table>
<thead>
<tr>
<th>Population size</th>
<th>Seed dictionary size</th>
<th>Fitness dictionary size</th>
<th>Crossover rate</th>
<th>Mutation rate</th>
<th>Setting Parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>512</td>
<td>50</td>
<td>3500</td>
<td>0.9</td>
<td>0.2</td>
<td></td>
</tr>
<tr>
<td>52.4</td>
<td>78.6</td>
<td>16</td>
<td>22</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>52.4</td>
<td>61.5</td>
<td>4</td>
<td>23</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>51.2</td>
<td>99.5</td>
<td>2</td>
<td>22</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>51.2</td>
<td>97.4</td>
<td>2</td>
<td>24</td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Performance results for seven cryptograms.
Multiple Sequence Alignment

Fundamental molecular biology problem:

• For "deciphering"
  – Identity common structure in a string of nucleotides (DNA, RNA) or amino acids (in proteins).
  – Phylogenetic relationships among organisms.
  – Evolutionary history

Very hard ~ on the order of \( m^n \) operations for \( n \) sequences of length \( m \).

Multiple Sequence Alignment
Clustal-W Example

From a set of DNA sequences such as these:

Clustal-W Example

Compute a phylogram such as this:
The Algorithm

- The chromosomes were candidate alignments, represented in a matrix with n rows.
- Create a random population of candidate alignments,
- For each candidate, apply variation operators to derive a child candidate.
- Apply a selection operator (fitness test) to generate the next generation.
- Repeat until 200 generations, no change in 100 generations, or number of gaps fell below a certain threshold.
Evolutionary Programming

Evolutionary Programming

Chellapilla and Fogel's EP is comparable to Clustal-W.

- Outperforms Clustal-W for low-similarity sequences.
- Tackles harder and longer sequences.
- Flexibility in fitness functions.

EP Advantages

<table>
<thead>
<tr>
<th>EP Score</th>
<th>7033</th>
<th>180</th>
<th>102</th>
<th>96</th>
<th>2744</th>
<th>3334 (3222, 346)</th>
<th>21</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4766</td>
<td>160</td>
<td>601</td>
<td>601</td>
<td>68.34</td>
<td>122.0 (122, 122)</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>7233</td>
<td>400</td>
<td>644</td>
<td>444</td>
<td>98.25</td>
<td>457.0 (475, 457)</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>2082</td>
<td>000</td>
<td>186</td>
<td>186</td>
<td>63.39</td>
<td>211.9 (211, 212)</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Data set used for testing the proposed evolutionary programming procedure for multiple sequence alignment. Information regarding the data set is provided in the Appendix.
Thank You!