Information Transmission Through Genetic Algorithm Fitness Maps

George Montañez

Machine Learning Department, Carnegie Mellon University, Pittsburgh, PA, USA

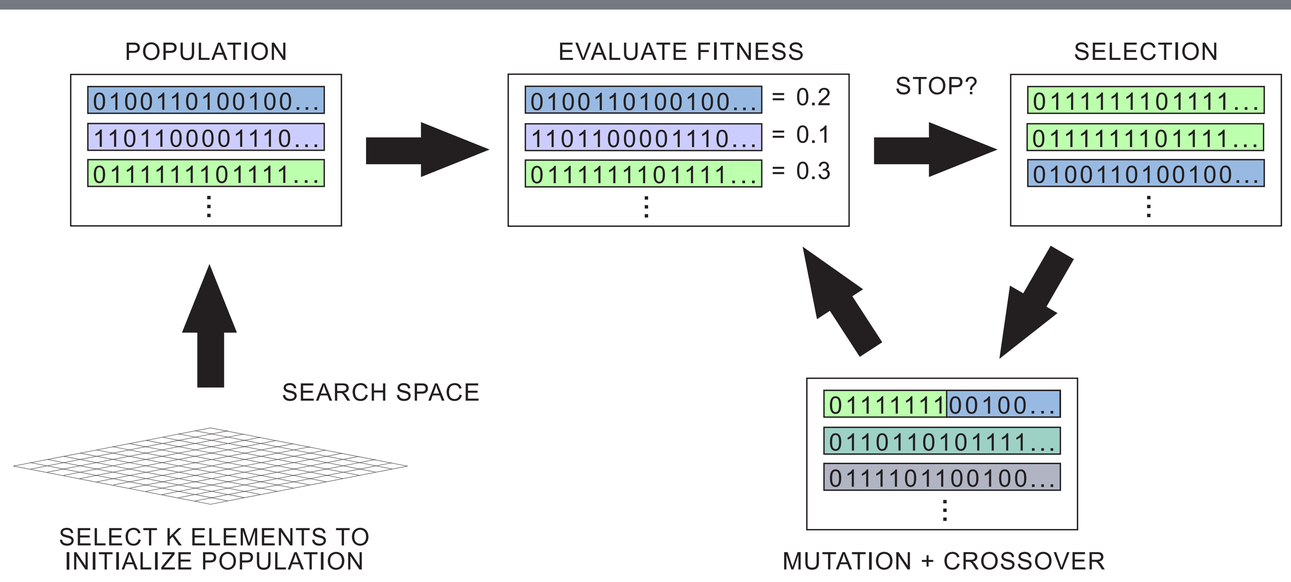
Abstract

To bound the amount of information transmitted from a fitness map to a genetic algorithm population, we use a method suggested by Abu-Mostafa *et al.* for measuring the information storage capacity of general forms of memory and represent the genetic algorithm as a communication channel. Our results show that a number of bits linear in the size of the search space can be stored in a fitness map, but on average only a logarithmic number of bits can be stored within a genetic algorithm population of bounded size and finite precision representation. Our results place an upper bound on the rate at which information can be transmitted through, or generated by and later extracted from, a genetic algorithm under fairly general conditions.

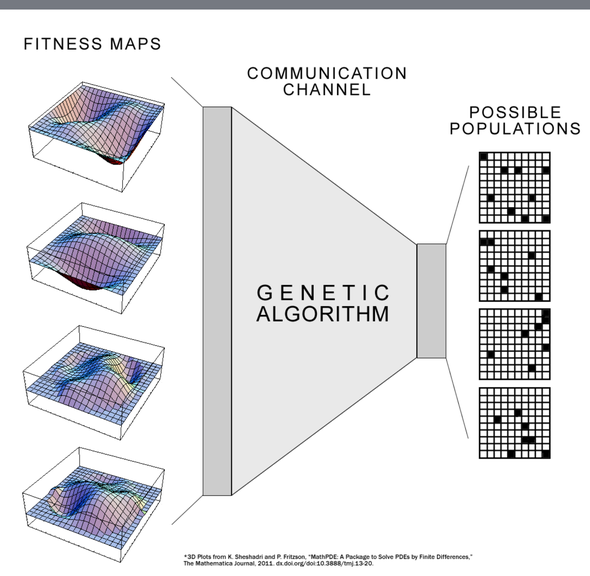
Introduction

- Evolution is a process by which species appropriate information from their environment.
- Genetic algorithms utilize key aspects of Darwinian evolution, including replication, mutation/crossover and selection.
- Genetic algorithms can generate only as much information as they can store in a population.
- We seek to determine the amount of information that can be sent through a genetic algorithm fitness map, and the rate of information transfer to a population.

The Control Flow Structure of a Typical Genetic Algorithm



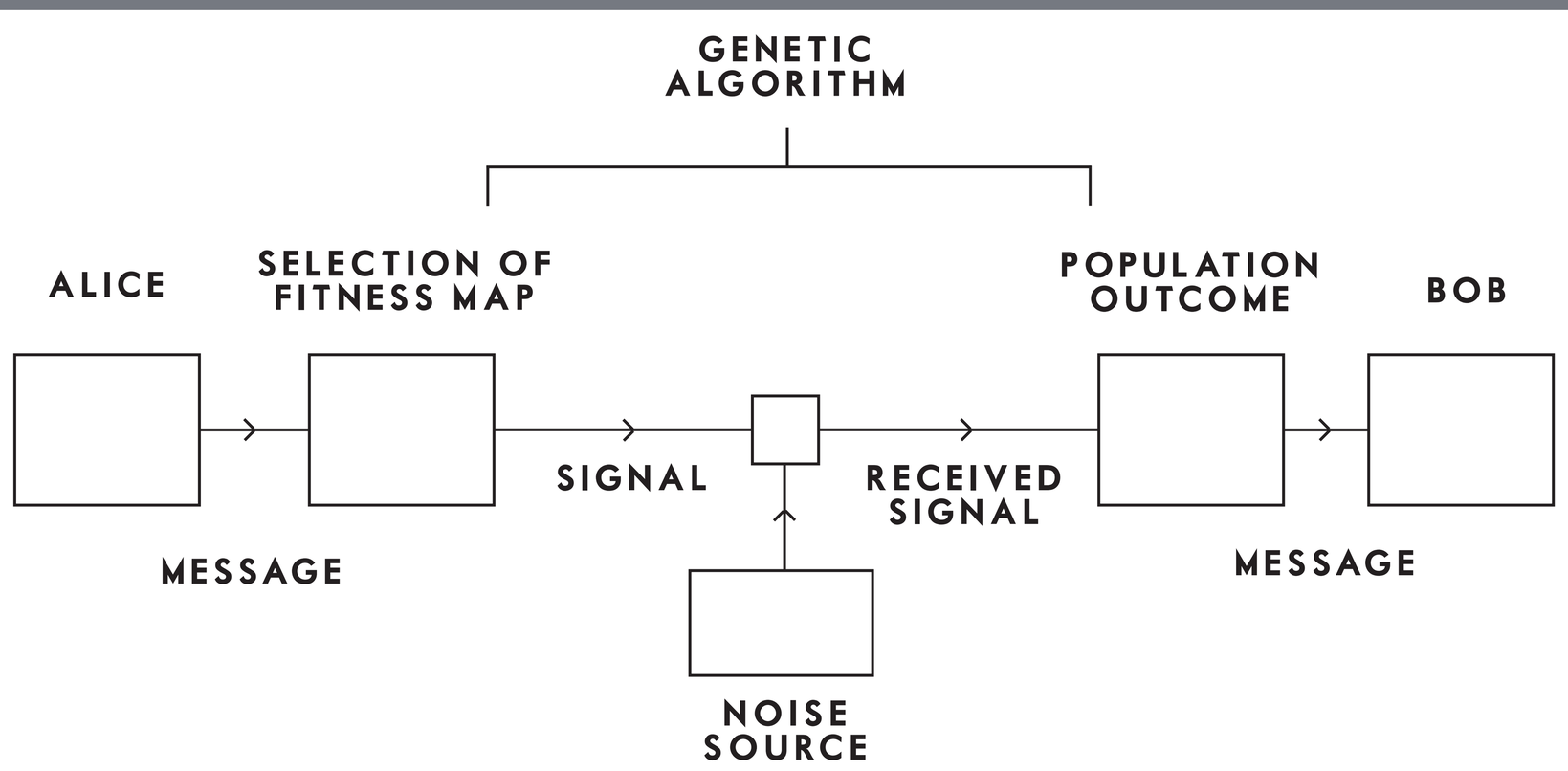
Messages and Outputs: GA Channel



Measuring Information Transmission

- Information flows from a fitness map to a population, as through a communication channel.
- Stochastic aspects of the genetic algorithm act as a noise source.
- Selecting a fitness map encodes information (message), transmitted through a genetic algorithm (noisy channel), and is output to a population (received message).
- Measuring the rate at which Alice can transmit to Bob through the channel upper bounds the rate at which information can flow through a genetic algorithm.

Genetic Algorithm as a Noisy Communication Channel



Would Anyone Ever Use a Genetic Algorithm as a Communication Channel?

- Short answer: No, not intentionally.
- However, it is possible to do so; therefore, the rules governing information flow through a communication channel apply.
- Researchers often select a particular fitness map (message), to produce a desired output (received message) with high probability.
  - Thus, we unwittingly use genetic algorithms as communication channels in practice.
  - Selecting specific fitness maps imparts (transmits) information.

Measuring Information Storage Capacity

Information Capacity C

- Count the number of outcomes a memory device can distinguish between and take the log base two of that number.
- Measures the maximum channel capacity, which occurs when each outcome is equally likely.
- For a finite search space $\mathcal{X}$ of possible organisms and populations having k members:

$$C = \log_2 \left((k + |\mathcal{X}| - 1) \text{ choose } k \right)$$

Key Results

For a finite search space $\mathcal{X}$:

- Discrete fitness maps have an information capacity C of $O(|\mathcal{X}|)$ bits.
- Output populations of bounded size have an information capacity C of $O(\log_2 |\mathcal{X}|)$ bits.
- The channel capacity (expected bits per population) is thus no more than $O(\log_2 |\mathcal{X}|)$.
- Genetic algorithms can output no more than $O(\log_2 |\mathcal{X}|)$ bits per population, on average.

Discussion

- Populations of bounded size cannot store more than a logarithmic (in the size of the search space) number of bits per population, on average.
- Therefore, genetic algorithms can generate no more than a logarithmic number of bits per population, on average.
- These bounds also apply to genetic algorithms using time-varying fitness functions.

Conclusion

- To measure how much information can be transmitted through a genetic algorithm from a fitness map to an output population, we represent the genetic algorithm as a noisy communication channel.
- We calculate the information capacity of both the input (fitness map) and the output (population) of that system.
- The information capacity of fitness maps is linear in the size of the search space: large storage capacity.
- The information capacity of population outcomes (for populations of bounded size) is logarithmic in the size of the search space.
- Therefore, the transmission process is lossy: on average, we can extract far less information from a genetic algorithm than we input through the selection of a fitness map.

Acknowledgements

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. 0946825 and the Ford Foundation Predoctoral Fellowship program. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author alone and do not necessarily reflect the views of the National Science Foundation, the Ford Foundation, Carnegie Mellon University or any other organization.

8003 Gates Hillman Complex - Machine Learning Department - Carnegie Mellon University - Pittsburgh, PA, USA Mail: gmontane@cs.cmu.edu WWW: http://ml.cmu.edu