APPENDIX

MODELING, SIMULATION AND APPLICATION OF BACTERIAL TRANSDUCTION IN GENETIC ALGORITHMS

Carlos Perales-Gravan, Javier de Vicente Buendia, Juan Castellanos, R. Lahoz-Beltra

NIT 2012. International Conference Natural Information Technologies. October 02-05, 2012, Madrid (Spain).
http://bioinformatica.net/tests/petri.html

Genetic Engineering Protocols in Evolutionary Computation

Transduction operator

In the genetic algorithm, the donors, Petri dish and bacterium, are selected by applying one of the following criteria. Let \( p^D \) be the donor Petri dish representing the sub-population from which to select the donor bacterium. The selection of the donor Petri dish is based upon one of the following criteria:

a) Random selection (method \( r \)).- In this case a Petri dish \( p^D \) is chosen at random from a uniform distribution according to the range \([1, P]\).

b) Maximum fitness (method \( \text{max} \)).- Given \( P \) Petri dishes, the Petri dish \( p^D \) with maximum fitness is selected. Thus, if \( f_i \) is the fitness value of the \( i \)th donor Petri dish, then we select \( \max_{1, \ldots, P} \{ f_1, f_2, \ldots, f_P \} \).

c) Average fitness (method \( \text{ave} \)).- Given \( P \) Petri dishes, the most representative Petri dish \( p^D \) with fitness value \( \bar{f} \) is selected, so that \( \min_{1, \ldots, P} |f_i - \bar{f}| \geq 0 \) where \( \bar{f} \) is the average fitness of the \( P \) donor dishes.

d) Roulette wheel dish selection (method \( \text{roul} \)).- In this case a Petri dish \( p^D \) is chosen by spinning a roulette wheel which assigns to each dish a slot whose arc size is proportional to its fitness value \( f_i \).

Once a donor Petri dish is selected, the choice of the donor bacterium \( b^D \) is made according to one of the following criteria:
a) Random selection (method $r$). - In this case a bacterium $b^D$ is chosen at random from a uniform distribution according to the range $[1, N]$.

b) Maximum fitness (method $max$). - Given $N$ bacteria, the bacterium $b^D$ with maximum fitness is selected. That is, so that $\max_{1 \leq i \leq N} \{ f_i, f_2, \ldots, f_j \}$ where $f_j$ is the fitness value of the $j$th donor bacterium.

c) Average fitness (method $ave$). - Given $N$ bacteria, the most representative bacterium $b^D$ with fitness value $f_j$ is selected, so that $\min_{1 \leq i \leq N} \left| f_j - \overline{f} \right| \geq 0$ where $\overline{f}$ is the average fitness of the bacterial population or donor Petri dish $p^D$.

d) Roulette wheel bacterial selection (method $roul$). - In this case a bacterium $b^D$ is chosen by spinning a roulette wheel which assigns to each dish a slot whose arc size is proportional to its fitness value $f_j$.

Note that this method is the well-known roulette wheel parent selection.

The recipient Petri dish and bacterium are both selected as follows: selection of the recipient Petri dish $p^R$ is conducted by means of one of the following criteria:

a) Random selection (method $r$). - In this case a Petri dish $p^R$ is chosen randomly from a uniform distribution according to the range $[1, P]$.

b) Minimum fitness (method $min$). - Given $P$ Petri dishes, the Petri dish $p^R$ with minimum fitness is selected. That is, $\min_{1 \leq i \leq P} \{ f_i, f_2, \ldots, f_i \}$ where $f_i$ is the fitness value of the $i$th recipient bacterium.

c) Average fitness (method $ave$). - Given $P$ Petri dishes, the most representative Petri dish $p^R$ with fitness value $\overline{f}_i$ is selected. Thus, $\min_{1 \leq i \leq P} \left| \overline{f}_i - \overline{f}_p \right| \geq 0$ where $\overline{f}_p$ is the average fitness of the $P$ recipient dishes.

d) Inverse roulette wheel dish selection (method $inv roul$). - In this case a Petri dish $p^R$ is chosen by spinning a roulette wheel which assigns to each dish a slot whose arc size is proportional to the inverse of its fitness value, thus $1/f_i$.

Once a recipient Petri dish is chosen, the selection of the receptor bacterium $b^R$ is conducted according to one of the following criteria:

a) Random selection (method $r$). - In this case a bacterium $b^R$ is chosen at random from a uniform distribution according to the range $[1, N]$.
b) Minimum fitness (method \( \text{min} \)). - Given \( N \) bacteria, the bacterium \( b^R \) with minimum fitness is selected. That is, \( \min_{1 \leq j \leq N} \{ f_1, f_2, \ldots, f_j \} \) where \( f_j \) is the fitness value of the \( j \)th recipient bacterium.

c) Average fitness (method \( \text{ave} \)). - Given \( N \) bacteria, the most representative bacterium \( b^R \) with fitness value \( f_j \) is selected, so that \( \min_{1 \leq j \leq N} | f_j - \bar{f} | \geq 0 \) where \( \bar{f} \) is the average fitness of the bacterial population or recipient Petri dish \( p^R \).

d) Inverse roulette wheel bacterial selection (method \( \text{inv roul} \)). - In this case a bacterium \( b^R \) is chosen by spinning a roulette wheel which assigns to each dish a slot whose arc size is proportional to the inverse of its fitness value, thus \( \frac{1}{f_j} \).

Transduction mechanism (bacterium DNA, white rectangle; bacteriophage DNA, grey rectangle). (a) Infection of a donor bacterium \( D \) with a bacteriophage. (b) Bacterial and bacteriophage DNA segments mix inside donor bacterium \( D \). A bacterial DNA segment is packed inside the bacteriophage ‘head’ (c), and is transferred to a recipient bacterium \( R \). Finally, inside the recipient bacterium, \( R \) homologous recombination or crossover occurs between the emigrant bacterial DNA segment and the target bacterial chromosome.
Figure a

Figure b
Transduction operator. (a) Transferring a chromosome segment: (step a) donor bacterium D (bacterium DNA, white rectangle), (step b) recipient bacterium R (bacterium DNA, grey rectangle) and chromosome segment D’ from donor bacterium, (step c) one chromosome results inside R following transduction. (b) Transferring a complete chromosome: (step a) donor bacterium D (bacterium DNA, white rectangle), (step b) recipient bacterium R (bacterium DNA, grey rectangle) and chromosome D’ from donor bacterium, (step c) one chromosome results inside R after transduction. (c) Migration and crossover: (step a) donor bacterium D (bacterium DNA, white rectangle), (step b) recipient bacterium R (bacterium DNA, grey rectangle) and the migrated chromosome D’ from donor bacterium, (step c) two chromosomes result inside R after migration and crossover.

Horizontal gene transfer (HGT) mechanisms in bacteria. Conjugation is a mechanism of HGT within population, whereas transduction is an HGT mechanism between populations.
Transduction experiment. The figure shows transduction from donor Petri dish ($p^D$) and bacterium ($b^D$) to recipient Petri dish ($p^R$) and bacterium ($b^R$). In the figure, $P$ is the total number of Petri dishes (or sub-populations), $N$ is the number of bacteria (or population size) per Petri dish and $r_e$ the number of experimental replicates.

Multiple bacterial colonies (or Petri dishes) are communicated via bacteriophages, cycling through generations searching for an optimum solution during $G$ generations.
AM radio receiver experiment with transduction transferring chromosome segments. Note how this kind of simulation experiment is an example of specialized transduction, where only three possible chromosome segments can be transferred by bacteriophages (for explanation see [Perales-Gravan and Lahoz-Beltra, 2008]).

SDS protocol (for details [Lahoz-Beltra, 2001]).
Conjugation operator

For an explanation see [Perales-Gravan and Lahoz-Beltra, 2008].

Bacterial conjugation operator. Once a pair of chromosomes $i$ and $j$ (or bacteria, D=donor, R=recipient) are selected from the same population (or Petri dish), the gene transfer occurs from a random point $i_o$ on the donor chromosome $i$ (D=chromosome origin). Since transfer of the donor chromosome is almost never complete, then the length $\ell$ of the strand (a copy) transferred to the recipient cell R is simulated with the Monte Carlo method, assuming DNA lengths exponentially distributed with a parameter $\alpha$ (conjugation parameter). Finally, the transferred strand experiences crossover, resulting a recombinant chromosome $j$ in bacterium R.

AM radio receiver experiment:

Figure a (left) b (right)
AM radio receiver. (a) Electronic circuit of a simple crystal radio showing the main electronic components to be optimized. (b) Radio enclosure details. (c) Bacterial chromosome with 14 genes codifying for the main characteristics of an AM radio receiver [from Perales-Gravan and Lahoz-Beltra, 2008].

**Statistical analysis**

The simulation experiments performance was evaluated as follows [Lahoz-Beltra and Perales-Gravan, 2010]. In first place, we obtained the average fitness of each Petri dish at the last generation $G$. Considering that $P$ is the total number of Petri dishes and $r_e$ the number of experimental replicates, then the total number of simulation trials was equal to $P \cdot r_e$. Subsequently, a Multiple Box-and-Whisker Plot [Tukey, 1977] was obtained with the average fitness values, considering that each plot represents an experimental protocol with $P \cdot r_e$ values. Note that we used a Notched Box-and-Whisker Plot. In this plot, a confidence interval for the median is provided by a notch surrounding the median. The endpoints of the notches are located at the median $\pm 1.5 \frac{IQR}{\sqrt{n}}$ so that the medians of two boxplots are significantly different at approximately the 0.05 level if the corresponding notches do not overlap. In each experimental protocol, the Notched Box-and-Whisker Plot indicates with a cross the mean of the average fitness values (MAF$_T$) for the $P \cdot r_e$ Petri dishes of an experiment. The MAF$_T$ value is given by the expression:

$$MAF_T = \frac{\sum_{k=1}^{P \cdot r_e} f_k}{P \cdot r_e}$$
Next, we performed a Kruskal-Wallis test evaluating the assumption that the medians of the average fitness values obtained under different protocols are equal. Note that experimental efficiency increases with an increase in the mean of the average fitness values (MAF), that is, when the MAF value displaces more to the top inside the box. The box represents the interquartile range of the average fitness values in the Box-and-Whisker Plot. Furthermore, since the length of the box representing the interquartile range is a measure of variability, given two experimental protocols, that protocol with the longest box will be the protocol driving the population to reach a higher number of optimum solutions [Perales-Gravan and Lahoz-Beltra, 2008].

Only in experiment 3 did we calculate the mean of the average fitness values, but now taking into account the Petri dish class. Thus, we obtained MAF values for the donor Petri dishes (MAF_D), the recipient Petri dishes (MAF_R) and for the dishes that were neither recipients nor donors (MAF_{DR}):

\[
\begin{align*}
    \text{MAF}_D &= \frac{\sum T_D}{D} \\
    \text{MAF}_R &= \frac{\sum T_R}{R} \\
    \text{MAF}_{DR} &= \frac{\sum_{k=1}^{T-D-R} T_k}{T-D-R}
\end{align*}
\]

where D is the number of donor Petri dishes and R the number of recipient Petri dishes.

Note that in all the optimization problems (thus, the simulation experiments) D=R, D and R being equal to 1. Therefore, we selected only one donor and recipient Petri dish per generation.

---

**Results: Figures**

In this section we include all the Figures cited in the paper.
Figure 3. Michalewicz function experiments with transduction transferring complete chromosomes. The figure shows the medians (notches) and MAF (crosses) of the average fitness values obtained in fifteen different simulation experiments: Conjugation with $p_t$ equal to 0% (without transduction) (1), 25% (2), 50% (3), 75% (4) and 100% (5). One-point recombination with $p_e$ equal to 0% (without transduction) (6), 25% (7), 50% (8), 75% (9) and 100% (10). Two-point recombination with $p_e$ equal to 0% (without transduction) (11), 25% (12), 50% (13), 75% (14) and 100% (15).

Figure 4. Knapsack problem experiments with transduction transferring complete chromosomes. The figure shows the medians (notches) and MAF (crosses) of the maximum fitness values obtained in fifteen different simulation experiments: Conjugation with $p_e$ equal to 0% (without transduction) (1), 25% (2), 50% (3), 75% (4) and 100% (5). One-point recombination with $p_e$ equal to 0% (without transduction) (6), 25% (7), 50% (8), 75% (9) and 100% (10). Two-point recombination with $p_e$ equal to 0% (without transduction) (11), 25% (12), 50% (13), 75% (14) and 100% (15).
Figure 5. Multiple Box-and-Whisker Plot for the AM radio receiver problem. The figure shows the medians (notches) and MAF (crosses) of the average fitness values obtained in thirty-three different simulation experiments (donor Petri dish - donor bacterium, recipient Petri dish - recipient bacterium). The following experiments involved transferring chromosome segments with $p_t$ equal to 100%: (1) Control experiment (without transduction). (2) r-r, r-r. (3) max-max, max-max. (4) ave-ave, ave-ave. (5) roul-roul, roul-roul. (6) min-min, min-min. (7) inv roul-inv roul, inv-roul-inv roul. (8) roul-roul, min-min. (9) roul-roul, ave-ave. (10) roul-roul, inv-roul-inv roul. (11) roul-roul, max-max. (12) roul-roul, r-r. (13) min-min, roul-roul. (14) ave-ave, roul-roul. (15) inv roul-inv roul, roul-roul. (16) max-max, roul-roul. (17) r-r, roul-roul. (18) max-max, roul-roul. (19) max-max, roul-r. (20) max-max, roul-min. (21) max-max, roul-ave. (22) max-max, roul-max. (23) max-max, roul-inv roul. (24) max-max, max-roul. (25) max-max, max-r. (26) max-max, max-min. (27) max-max, max-ave. (28) max-max, max-inv roul. Experiments transferring chromosome segments with $p_t$ equal to: (29) 75% and max-max, roul-r. (30) 50% and max-max, roul-r. (31) 25% and max-max, roul-r. (32) 0% and max-max, roul-r. Finally, experiment (33) involved transferring a complete chromosome with $p_t$ equal to 100% and max-max, roul-r.
Figure 6. Multiple Box-and-Whisker Plot for the AM radio receiver problem. The figure shows the medians (notches) and MAF (crosses) of the average fitness values obtained in the (a) control experiment (without transduction) and the two best transduction experiments: (b) max-max, ru-l-r (transferring chromosome segments). (c) max-max, ru-l-r (transferring a complete chromosome).

Figure 7a
Figure 7. Representative performance graph for the AM radio receiver problem showing the Box-and-Whisker Plot per generation: (a) control experiment (without transduction) and the best two transduction experiments: (b) max-max, rul-r (transferring chromosome segments), (c) max-max, rul-r (transferring a complete chromosome).
Figure 8. Performance graph for the AM radio receiver problem: (a) control experiment (without transduction) showing the mean of the average fitness per generation for the total of Petri dishes (MAF). The performance graph in the best two transduction experiments: (b) max-max, nul-r (transferring chromosome segments) and (c) max-max, nul-r (transferring a complete chromosome) showing MAF (thick gray line), MAF (dashed line), MAF (peaks and valleys line) and MAF (solid line) values.

Bibliography

[Lahoz-Beltra, 2001] R. Lahoz-Beltra. 2001. Evolving hardware as model of enzyme evolution. BioSystems 61: 15-25.
[Lahoz-Beltra and Perales-Gravan, 2010] R. Lahoz-Beltra, C. Perales-Gravan. 2010. A survey of nonparametric tests for the statistical analysis of evolutionary computational experiments. International Journal Information Theories and Applications 17(1): 49-61.
[Perales-Gravan and Lahoz-Beltra, 2008] C. Perales-Gravan, R. Lahoz-Beltra. 2008. An AM radio receiver designed with a genetic algorithm based on a bacterial conjugation genetic operator. IEEE Transactions on Evolutionary Computation 12(2): 129-142.
[Tukey, 1977] J.W. Tukey. 1977. Exploratory Data Analysis. Addison-Wesley, Reading, MA.

Corresponding author

R. Lahoz-Beltra: Dept. of Applied Mathematics, Faculty of Biological Sciences, Complutense University of Madrid, 28040 Madrid, Spain; e-mail: lahozraf@bio.ucm.es