Generalized Net Model of Selection Function Choice in Genetic Algorithms

Tania Pencheva[1], Krassimir Atanassov[1], Anthony Shannon[2]

[1] Institute of Biophysics and Biomedical Engineering, Bulgarian Academy of Sciences
105 Acad. G. Bonchev Str., Sofia-1113, Bulgaria
tania.pencheva@clbme.bas.bg , krat@bas.bg
[2] Faculty of Engineering & IT, University of Technology
Sydney, NSW 2007, Australia
tshannon38@gmail.com, Anthony.Shannon@uts.edu.au

Abstract

The apparatus of Generalized Nets is applied here to describe the user's choice of a selection function serves as a basis for one of the basic genetic algorithm operators, namely the *selection operator*. The selection is a probabilistic process based upon the individual's fitness such that the better individuals have an increased chance of being selected for the next generation. Selection of best individuals could be realized using different methods. In the present study the user is allowed to choose between two of the most widespread selection functions, namely *roulette wheel selection* and *stochastic universal sampling*. The resulting generalized net model could be considered as a separate module, but it can also be assembled into a generalized net model to describe a whole genetic algorithm.

Keywords: Generalized nets, Genetic algorithms, Selection function.

1 Introduction

Genetic Algorithms (GA) are an adaptive heuristic search algorithm, designed to simulate processes in natural systems necessary for evolution, and especially those that follow the principles of "survival of the fittest" formulated for first time by Charles Darwin [6]. GA are implemented in a computer simulation in which a population of abstract representations (called *chromosomes* or the *genotype of the genome*) of candidate solutions (called *individuals*, *creatures*, or *phenotypes*) to an optimization problem evolves toward better solutions. Once the genetic representation and the fitness function are defined, GA proceed to initialize a population of solutions *randomly*, then improve it through repetitive application of *mutation*, *crossover*, *inversion* and *selection* operators. In each generation, the *fitness* of every individual in the population is evaluated, multiple individuals are stochastically selected from the current population (based on their fitness), and modified (recombined and possibly randomly mutated) to form a new population. The new population is then used in the next iteration of the algorithm. Commonly, the algorithm terminates when either a maximum number of generations has been produced, or a satisfactory fitness level has been reached for the population.

Due to a variety of successive implementations of Generalized Nets (GN) theory for description of parallel processes in several areas [1-3], the idea of using GN for the description of GA has intuitively appeared. Up to now, a few GN models regarding GA performance have been developed [1, 3, 9-14]. A GN model for GA learning is proposed in [1, 3]. In [9-12] GN models are used to describe the basic genetic algorithms operators, correspondingly *selection*, *crossover* and *mutation* have been developed. The GN model of a *roulette wheel selection* method, which is one of the widely used selection functions, has been developed in [9], while the GN model of a *stochastic universal sampling* is presented in [10]. Different types of *crossover*, namely *one-*, *two-point crossover*, as well as "*cut and splice*" techniques, are described in details in [11]. The GN model, presented in [12], describes the *mutation* operator of the Breeder GA. The evaluation of GA fitness function is presented by the GN model developed in [13], while the selection of GA operators is described by the GN model, presented in [14].

The purpose of the present investigation is to develop a GN model, which allows the user to choose a selection method that to be implemented in the *selection* operator. This GN model combines the selection methods of *roulette wheel selection* [9] and the *stochastic universal sampling* [10] as the mostly used, but could be expanded with a number of other possible methods.

2 Selection Methods

The selection of individuals to produce successive generations plays an extremely important role in a genetic algorithm. A probabilistic selection is performed based upon the individual's fitness such that the better individuals have an increased chance of being. Most functions realizing the selection are stochastic and designed so that a small proportion of less fit solutions are selected. This helps keep the diversity of the population large, preventing premature convergence on poor solutions. There are many methods for selection of the best individuals, i.e. *roulette wheel selection*, *Boltzman selection*, *tournament selection*, *rank selection*, *steady state selection* and some others. Among the most popular and well-studied selection methods are ro*ulette wheel selection* and *stochastic universal sampling*. The selection method is a user-defined parameter of the entire procedure of selection.

2.1 Roulette Wheel Selection (RWS)

A common selection approach assigns a probability of selection P_j to each individual j based on its fitness value. A series of N random numbers is generated and compared against the cumulative probability

$$C_i = \sum_{j=1}^{i} P_i$$

of the population. The appropriate individual i is selected and copied into the new population, if $C_{i\text{-}1} < U(0,\ 1) \le C_i$. Various methods exist to assign probabilities to individuals: *roulette wheel*, *linear ranking* and *geometric ranking*.

Roulette wheel, developed by Holland [7], is the first selection method. The probability P_i for each individual is defined by:

$$P\ [\text{Individual } i \text{ is chosen}] = \frac{F_i}{\sum_{j=1}^{PopSize} F_j}$$

where F_i equals the fitness of individual i. The use of *roulette wheel selection* limits the genetic algorithm to maximization since the evaluation function must map the solutions to a fully ordered set of values on $\Re^+$. Extensions, such as windowing and scaling, have been proposed to allow for minimization and negativity.

In *roulette wheel selection* the individuals are mapped to contiguous segments of a line, such that each individual's segment is equal in size to its fitness. A random number is generated and the individual whose segment spans the random number is selected. The process is repeated until the desired number of individuals is obtained (called *mating population*). This technique is analogous to a *roulette wheel* with each slice being proportionally sized to the fitness

2.2 Stochastic Universal Sampling (SUS)

Stochastic universal sampling developed by Baker [4] is a single-phase sampling algorithm with minimum spread and zero bias. Instead of a single selection pointer employed in *roulette wheel* methods, *SUS* uses N equally spaced pointers, where N is the number of selections required. The population is shuffled randomly and a single random number *pointer1* in the range [0, $1/N$] is generated. The N individuals are then chosen by generating the N pointers, starting with *pointer1* and spaced by $1/N$, and selecting the individuals whose fitness spans the positions of the pointers. If $et(i)$ is the expected number of trials of individual i, $\lfloor et(i) \rfloor$ is the floor of $et(i)$ and $\lceil et(i) \rceil$ is the ceiling, an individual is thus guaranteed to be selected a minimum of times $\lfloor et(i) \rfloor$ and no more than $\lceil et(i) \rceil$, thus achieving minimum spread. In addition, as individuals are selected entirely on their positions in the population, *SUS* has zero bias. For these reasons, *SUS* has become one of the most widely used selection algorithms in current GA.

Fig. 1 demonstrates the *stochastic universal sampling*. The individuals are mapped to contiguous segments of a line, such that each individual's segment is equal in size to its fitness exactly as in *roulette wheel selection*. As many equally spaced pointers are placed over the line, as are the individuals to be selected (N). For 6 individuals (N = 6) to be selected, the distance between the pointers is 1/6 = 0.167. Fig. 1 shows the selection for the sample of the random number 0.1 in the range [0, 0.167].

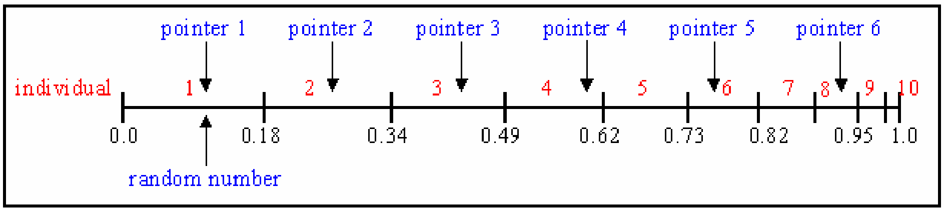

Fig. 1. *Stochastic universal sampling*

After selection the mating population consists of the individuals 1, 2, 3, 4, 6 and 8. *Stochastic universal sampling* ensures a selection of offspring which is closer to what is deserved than *roulette wheel selection*.

3 GN Model for Choosing of a Selection Method

GN model that allows the user to make a choice of the selection method is presented in Fig. 2. This model is built upon the GN models the separately describe the functions of *roulette wheel selection* method [9] and *stochastic universal sampling* [10] as presented in *Genetic Algorithms Toolbox* in Matlab [5, 8].

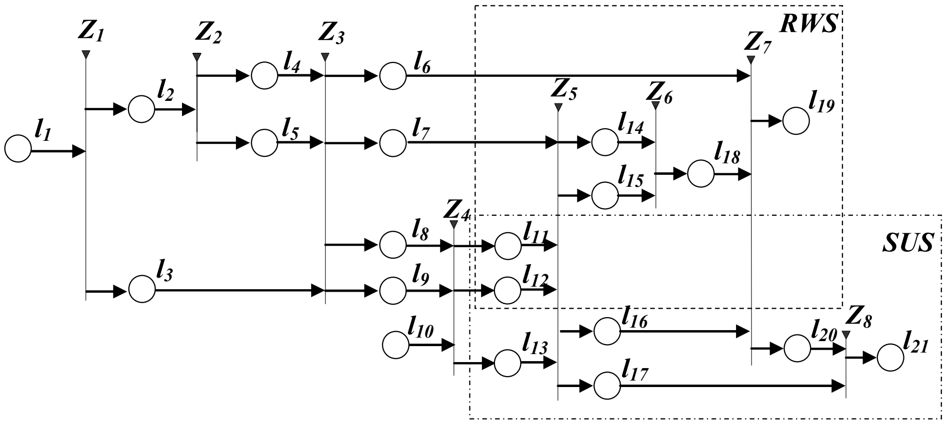

Fig. 2. GN model of selection function choice

The token α enters GN in place l_1 with an initial characteristic "pool of possible parents". The token α is split into new tokens β and γ, which correspondingly obtain characteristics "fitness values of the individuals in the population (*FitnV*)" in place l_2 and "number of individuals to be selected (*Nsel*)" in place l_3. The form of the first transition of the GN model is as follows:

$$Z_1 = <\{l_1\},\ \{l_2, l_3\}, r_1, \wedge(l_1)>$$

$$r_1 = \begin{array}{c|cc} & l_2 & l_3 \\ \hline l_1 & true & true \end{array}$$

The token β is split into new tokens δ and ε, which obtain correspondingly characteristics "calculation of the function *cumfit = cumsum(FitnV)*" in place l_4 and "identify the population size (*Nind*)" in place l_5. The form of the second transition of the GN model is as follows:

$$Z_2 = <\{l_2\},\ \{l_4, l_5\}, r_2, \wedge(l_2)>,$$

$$r_2 = \begin{array}{c|cc} & l_4 & l_5 \\ \hline l_2 & true & true \end{array}$$

Further, the tokens δ and γ are combined in a new token φ in place l_6 with a characteristic "calculation of the function *Mf = cumfit(:, ones(1, Nsel))*". The token ε keeps its characteristic "identify the population size (*Nind*)" in place l_7. The tokens δ and ε are combined in a new token η in place l_8 with a characteristic "calculation of the function *cumfit(Nind)"*. The token γ keeps its characteristic "number of individuals to be selected (*Nsel*)" in place l_9. The form of the third transition of the GN model is as follows:

$$Z_3 = <\{l_3, l_4, l_5\},\ \{l_6, l_7, l_8, l_9\}, r_3, \wedge(l_3, l_4, l_5)>$$

$$r_3 = \begin{array}{c|cccc} & l_6 & l_7 & l_8 & l_9 \\ \hline l_3 & true & false & false & true \\ l_4 & true & false & true & false \\ l_5 & false & true & true & false \end{array}$$

A new token ν with a characteristic "choice of selection function – *RWS* or *SUS*" enters GN in place l_{10}. The token η keeps its characteristic "calculation of the function *cumfit(Nind)*" in place l_{11}. The token γ obtains a new characteristic "*rand*(*Nsel*)" in place l_{12}. The tokens η and γ are combined in a new token λ'' (the notation ' is used for *RWS* and the notation '' – for *SUS*) in place l_{13} with a characteristic "calculation of the function

*trials = cumfit(Nind) / Nsel * (rand + (0:Nsel-1)')*".

The form of the fourth transition of the GN model is as follows:

$$Z_4 = <\{l_8, l_9, l_{10}\},\ \{l_{11}, l_{12}\}, r_4, \wedge(l_8, l_9, l_{10})>$$

$$r_4 = \begin{array}{c|ccc} & l_{11} & l_{12} & l_{13} \\ \hline l_8 & true & false & W_{8,13} \\ l_9 & false & true & W_{9,13} \\ l_{10} & true & true & W_{10,13} \end{array}$$

where $W_{8,13} = W_{9,13} = W_{10,13}$ = "*SUS* is chosen".

The token ε keeps its characteristic "identify the population size (*Nind*)" in place l_{14}. The tokens η and γ are combined in a new token λ' in place l_{15} with a characteristic "calculation of the function

*trials = cumfit(Nind) .*rand(Nsel, 1)*".

The tokens ε and λ'' are combined in a new token θ'' in place l_{16} with a characteristic "calculation of the function *Mt = trials(:, ones(1, Nind))'* ". The token γ obtains a new characteristic "*sort*(*rand*(*Nsel*))" in place l_{17}. The form of the fifth transition of the GN model is as follows:

$$Z_5 = <\{l_7, l_{11}, l_{12}, l_{13}\},\ \{l_{14}, l_{15}, l_{16}, l_{17}\}, r_5, \wedge(l_7, l_{11}, \vee(l_{12}, l_{13}))>$$

$$r_5 = \begin{array}{c|cccc} & l_{14} & l_{15} & l_{16} & l_{17} \\ \hline l_7 & true & false & W_{7,16} & false \\ l_{11} & false & W_{11,15} & false & false \\ l_{12} & false & W_{12,15} & false & W_{12,17} \\ l_{13} & false & false & W_{13,16} & false \end{array}$$

where $W_{11,15} = W_{12,15}$ = "*RWS* is chosen" and $W_{7,16} = W_{11,16}$ = "there is a token in place l_{13}".

The tokens ε and λ' are combined in a new token θ' in place l_{18} with a characteristic "calculation of the function *Mt = trials(:, ones(1, Nind))'* ". The form of the sixth transition of the GN model is as follows:

$$Z_6 = <\{l_{14}, l_{15}\}, \{l_{18}\}, r_6, \wedge(l_{14}, l_{15}) >$$

$$r_6 = \begin{array}{c|c} & l_{18} \\ \hline l_{14} & true \\ l_{15} & W_{15,17} \end{array}$$

where $W_{15,17}$ = "there is a token in place l_{15}".

The tokens φ and θ' are combined in a new token ω' in place l_{19} with a characteristic "calculation of the function:

[NewChrIx, ans] = find (Mt < Mf & ...[zeros(1, Nsel); Mf(1:Nind-1, :)] ≤ Mt)".

The tokens φ and θ'' are combined in a new token ω'' in place l_{20} with a characteristic "calculation of the function:

[NewChrIx, ans] = find (Mt < Mf & [zeros(1, Nsel); Mf(1:Nind-1, :)] ≤ Mt)".

The form of the seventh transition of the GN model is as follows:

$$Z_7 = <\{l_6, l_{16}, l_{18}\}, \{l_{19}, l_{20}\}, r_7, \wedge(l_6, \vee(l_{16}, l_{18})) >$$

$$r_7 = \begin{array}{c|cc} & l_{19} & l_{20} \\ \hline l_6 & true & true \\ l_{16} & false & W_{16,20} \\ l_{18} & W_{18,19} & false \end{array}$$

where $W_{16,20}$ = "there is a token in place l_{16}" and $W_{18,19}$ = "there is a token in place l_{18}".

In the place l_{19} the new chromosome is created and the selection function, performing *roulette wheel selection* method, is completely fulfilled.

The tokens ω'' and γ are combined in a new token σ'' in place l_{21} with a characteristic "shuffle new population *NewChrIx = NewChrIx(shuf)*". The form of the eighth transition of the GN model is as follows:

$$Z_8 = \langle \{l_{17}, l_{20}\}, \{l_{21}\}, r_8, \wedge(l_{17}, l_{20}) \rangle$$

$$r_8 = \begin{array}{c|c} & l_{21} \\ \hline l_{17} & true \\ l_{20} & W_{20,21} \end{array}$$

where $W_{20,21}$ = "there is a token in place l_{20}".
In place l_{21} the new chromosome is created and the selection function, performing *stochastic universal sampling*, is completely fulfilled.

4 Analysis and Conclusions

The theory of Generalized Nets has been applied here in order to allow the user to choose a selection function to serve as the basis of one of the genetic algorithm operators, namely the *selection operator*. The GN model developed in this paper permits the user to choose between two of the mostly widespread selection functions, namely *roulette wheel selection* and *stochastic universal sampling*, but it can be expanded with other selection methods too. Such a GN model could be considered as a separate module, but it can also be assembled into a single GN model for the description of a whole genetic algorithm.

Acknowledgments

This work is supported by the National Science Fund of Bulgaria by Grant DID-02-29 "Modeling Processes with Fixed Development Rules (ModProFix)".

References

[1] Aladjov H., K. Atanassov, A Generalized Net for Genetic Algorithms Learning, *Proceedings of the XXX Spring Conference of the Union of Bulgarian Mathematicians*, Borovets, Bulgaria, April 8-11, 2001, 242-248.

[2] Atanassov K., *Generalized Nets*, World Scientific, Singapore, New Jersey, London, 1991.

[3] Atanassov K., H. Aladjov, *Generalized Nets in Artificial Intelligence, Vol. 2: Generalized Nets and Machine Learning*, Prof. M. Drinov

Academic Publishing House, Sofia, 2000.
[4] Baker J., Reducing Bias and Inefficiency in the Selection Algorithm, *Proceedings of the Second International Conference on Genetic Algorithms and their Application*, Hillsdale, New Jersey, 1987, 14-21.
[5] Chipperfield A., P. J. Fleming, H. Pohlheim, C. M. Fonseca, *Genetic Algorithm Toolbox for Use with MATLAB*, 1993.
[6] Goldberg D. E., *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison Wesley, 1989.
[7] Holland J., *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann Arbor, 1975.
[8] MathWorks, *Genetic Algorithm Toolbox User's Guide for MATLAB*.
[9] Pencheva T., K. Atanassov, A. Shannon, Modelling of a Roulette Wheel Selection Operator in Genetic Algorithms Using Generalized Nets, *Bioautomation*, 2009, 13(4), 257-264.
[10] Pencheva T., K. Atanassov, A. Shannon, Modelling of a Stochastic Universal Sampling Selection Operator in Genetic Algorithms Using Generalized Nets, *Proceedings of the Tenth International Workshop on Generalized Nets*, December 5, 2009, Sofia, 1-7.
[11] Pencheva T., O. Roeva, A. Shannon, Generalized Net Models of Crossover Operator of Genetic Algorithm, *Proceedings of Ninth International Workshop on Generalized Nets*, Sofia, Bulgaria, July 4, 2008, 2, 64-70.
[12] Roeva O., A. Shannon, A Generalized Net Model of Mutation Operator of the Breeder Genetic Algorithm, *Proceedings of Ninth International Workshop on Generalized Nets*, Sofia, Bulgaria, July 4, 2008, 2, 59-63.
[13] Roeva O., K. Atanassov, A. Shannon, Generalized Net for Evaluation of the Genetic Algorithm Fitness Function, *Proceedings of the Eighth International Workshop on Generalized Nets*, Sofia, Bulgaria, June 26, 2007, 48-55.
[14] Roeva O., K. Atanassov, A. Shannon, Generalized Net for Selection of Genetic Algorithm Operators, *Annual of "Informatics" Section*, *Union of Scientists in Bulgaria*, 2008, 1, 117-126.