Comparing Grammatical Evolution Survivor Selection Methods in Forecasting Tidal Level

Nerfita Nikentari*, M Radzi Rathomi and Bondan Chorisma
Departement of Informatics, Faculty of Engineering, Universitas Maritim Raja Ali Haji, Tanjungpinang, Indonesia

* nerfita.nikentari@umrah.ac.id

Abstract. Grammatical Evolution (GE) is a grammar-based form of Genetic Programming. One important operator in GE algorithms is the selection scheme. The objective of this study is a comparison of two models of survivor selection, the generalation replacement and steady state, for use in tidal level forecasting. The survivor selection schemes are compared and evaluated according to their properties. The results show the performance of steady state is 95.66% and 94.39% for generalation replacement.

1. Introduction
GE is a branch of evolutionary computation based on grammar. GE is a grammar-based form of Genetic Programming (GP). Formal grammar is used in the genotype to phenotype mapping process in GE [1][2]. The foundation of the GE system is the genotype-phenotype mapping process used to generate the rule-set. Each genotype, represented as a variable length binary string, contains in its codons[3]. A grammar that used to describe the structures that are generated by GE typically represented in plain text Backus-Naur Form (BNF) format[4]. Each individual in GE is internally represented as an array of integers, a genome. GE uses a generative grammar to guide the construction of a phenotype string output from this genome input.

A BNF grammar is defined as 4-tuple \{N, T, P, S\}, where
- N is the set of non-terminals,
- T is the set of terminals,
- P is a set of production rules that map the elements of N to T, and
- S is a start symbol

GE algorithm is similar to a standard genetic algorithm (GA). The purpose of this research is to compare survivor selection method in GE. Comparison will show which survivor selection is better in forecasting tidal level.

2. Related Work
Smith and Vavak[5] compared a number of selection and replacement strategies for use in Steady State Genetic Algorithms. Genetic Algorithms using Steady State models showed a better result to track moving optima than using Generational models, but implementing the former requires an additional choice of which members of the current population should be replaced by new offspring.
Blickle and Thiele [6] analyze and compare several selection schemes used in GA, namely tournament selection truncation selection and linear and exponential ranking selection and fitness proportional selection. Compared and contrasted behavior between generational and steady-state used genetic algorithms has studied [7]. Feature that used for comparing the behaviour are replacement step each time, selection criterion for copying, selection criterion for deleting, relation to schema theorem, population diversity, ideal performance, new elite availability in subsequent steps, and application fields. Nhita et. al [8] compared performance of Grammatical Evolution and Adaptive Neuro-Fuzzy Inference System (ANFIS) to forecast rainfall in Bandung. Pagan et. al [9] shows the robustness of the Grammatical Evolution algorithm for migraine prediction against sensor failures. This research shows that Grammatical Evolution computes automatically feature selection and obtains simple non-linear mathematical expressions for the prediction models. Hilburg et. al [10] proposed modeling approach using Gramatical Evolution to model the dynamic power consumption of enterprise servers in data centers.

3. Data and Methodology

Data in this research is a dataset of tidal level that recorded every 1 hour. The data is transformed to daily data using average since the purpose is to making daily forecasting. Data will be divided into two parts, training and testing data. A transformation performed on input to ensure the data distributed evenly and scale it into an comparable and acceptable range. This transformation is called data normalization.

Data accuracy is measured using mean absolute percentage error (MAPE). MAPE used to compare forecast accuracy on several series with different scales[11], if the time series have zero values, the MAPE yields undefined results because of a division by zero[12]. Implementation of Grammatical Evolution used in the experiments are:

1. Initialization

   Initialization is the initial stage performed, in which the population will be raised randomly through a procedure. Each chromosome contains the genes generated by a positive integer. Each chromosome has a different gene length, the maximum length of the previous gene should be determined first. Thus, in one population, the length of the chromosome that can form is in the range [1, the maximum number of genes].

2. Chromosome Decode

   Chromosome decode is used to convert each chromosome into a generation into a rule by using the production rules that are obscured on BNF. Each chromosome that does not produce a function, will be duplicated. Duplication is done by copying the same chromosome to a random sequence, otherwise if a chromosome has produced a function before the gene in the chromosome is consumed pruning, the cropped gene will not be used[13].

3. Evaluation

   Individual evaluation is performed only when a chromosome has produced a function derived from the chromosome decoding process, each result of the obtained function will be searched the difference between the actual data and the predicted data to be used to find the fitness value

4. Elitism

   Elitism main purpose is to copy the chromosome with the highest fitness value of each population on a temporary basis. If there is a chromosome that has the lowest fitness value in the next generation, then the copied chromosome will replace the chromosome and if in the next generation there is a chromosome that has a higher fitness value than the previously copied chromosome then the previous chromosome position will be replaced.
5. Selection
Selection is a process of choosing a chromosome as a parent. Parent selection is done by generating a random value between 0 and 1. If the random value is smaller than the chromosome value, then the chromosome is selected as the parent.

6. Crossover
Crossover is a process combining two or more parent chromosomes for generating a new offspring. Generating can be done by probability decision. In this research we applied two point crossover.

7. Mutation
Mutation is the process of converting one or more genes into a single chromosome, based on the probability of mutation. This mutation process aims to change the content of a gene in a chromosome.

8. Survivor Selection
In this process all new chromosomes in the crossover and mutation process are replacing the entire old chromosome regardless of age and its fitness value. This research comparing generalation replacement and steady state for survivor selection.

4. Experimental Scenario and Result
The testing is conducted towards the use of BNF and the combination of evolution parameter. BNF is defined as the characteristics of data recasting of time series. BNF grammar used for tide level prediction shows in table 1.

| Table 1. BNF Grammar |
|----------------------|
| N = \{expr, op, const, var_x, var_y\} |
| T = \{+, -, *, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 1.1, 1.2, 1.3, 1.4, x1, x2, x3, x4, x5, x6, x7, y1, y2, y3, y4, y5, y6, y7\} |
| S = \{expr\} |
| P, formed by the following rules of production : |
| 1. <expr> := (<const><op><var_x>) <op> (<const><op><expr>) (A) |
| | (<const><op><var_x>)<op>(<const><op><var_y>) (B) |
| | (<const><op><var_x>)<op><var_y> (C) |
| 2. <op> := + (A) |
| | - (B) |
| | * (C) |
| 3. <const> := 0, 1 (A) |
| | 0, 2 (B) |
| | 0, 3 (C) |
| | 0, 4 (D) |
| | 0, 5 (E) |
| | 0, 6 (F) |
| | 0, 7 (G) |
| | 0, 8 (H) |
| | 0, 9 (I) |
| | 1 (J) |
| | 1, 1 (K) |
| | 1, 2 (L) |
| | 1, 3 (M) |
| | 1, 4 (N) |
| 4. <var_x> := x1 (A) |
| | x2 (B) |
The function based on BNF grammar as shown in table 1 is:

$$(1*x_1) + (1.3+(1.3*x_3)+(1.3+(0.2+x_6)+0.3)$$  \hspace{1cm} (1)$$

BNF shows in table 1 is using for testing several scenarios for generational replacement and steady state survivor selection. In this phase different kind values of GE parameter such as generation size, population size, crossover probability, mutation probability. The experiment scenario using these four parameters shows in Table 2.

| Table 2. Grammatical Evolution Parameters |
|------------------------------------------|
| Generation Size | Population Size | Crossover Probability | Mutation Probability |
|-----------------|-----------------|-----------------------|----------------------|
| 30              | 80              | 0.7                   | 0.2                  |
|                 |                 | 0.3                   |                      |
|                 |                 | 0.3                   |                      |
| 100             | 80              | 0.7                   | 0.2                  |
|                 |                 | 0.3                   |                      |
|                 |                 | 0.3                   |                      |
| 50              | 80              | 0.7                   | 0.2                  |
|                 |                 | 0.3                   |                      |
|                 |                 | 0.3                   |                      |
| 100             | 80              | 0.7                   | 0.2                  |
|                 |                 | 0.3                   |                      |
|                 |                 | 0.3                   |                      |

Multiple experiments shows the best parameter values for generational replacement and steady state survivor selection shows in table 3

| Table 3. Best Parameter Values for generational replacement and steady state |
|------------------------------|-----------------|-----------------|
|                              | Generational Replacement | Steady State |
| Generation Size              | 50               | 30              |
| Population Size              | 100              | 100             |
| Crossover Probabilty         | 0.8              | 0.7             |
Comparison between actual data and forecast data using generalation replacement and steady state can be seen in fig 1 and fig 2.

![Figure 1. Comparison of actual data vs forecast data with steady state](image1.png)

![Figure 2. Comparison of actual data vs forecast data with generalation replacement](image2.png)

### 5. Conclusion

Two survival selection method, generational peplacement and steady state, was applied in this paper to forecast tidal level. Several paremeters (generation size, population size, crossover probabilty, mutation probability) with different values showed that the combination of crossover probabilities gives a significant error on each survivor selection method. The experimental results obtained the best results on GE steady state is 95.66 % while the performance of generational peplacement obtained result is 94.39 %.

Better accuracy in forecasting can provide accurate information about tidal level. In further research, to get better result in forecasting with GE, not only using one algorithm for selection process. There are many selection algorithm can be used for select the best chromosomes besides roulette wheel, for example linear ranking selection, trunction selection, Boltzman selection, tournament selection and rank selection.

### References

[1] J. R. Koza. Genetic Programming: On The Programming of Computers by Means of Natural Selection, Volume 1. MIT press, (1992).

[2] M. O'Neill and C. Ryan. Grammatical Evolution: Evolutionary Automatic Programming in a Aarbitrary Language, (2003).
[3] Brabazon, A., et al. “Grammar-Mediated Time-Series Prediction.” Journal of Intelligent Systems, vol. (2005), no. 2-3

[4] Fenton, M., Mcdermott, J., Fagan, D., Forstenlechner, S., Hemberg, E., & Oneill, M, “PonyGE2,”. Proceedings of the Genetic and Evolutionary Computation Conference Companion on - GECCO 17, (2017).

[5] Smith, J. E., & Vavak, F. Replacement Strategies in Steady State Genetic Algorithms: Dynamic Environments. Journal of Computing and Information Technology, (1999), 49-59.

[6] Blickle, T., & Thiele, L. A Comparison of Selection Schemes Used in Evolutionary Algorithms. Evolutionary Computation, 4(1996), 361-394.

[7] Vavak, F., & Fogarty, T. (n.d.). Comparison of steady state and generational genetic algorithms for use in nonstationary environments. Proceedings of IEEE International Conference on Evolutionary Computation.

[8] F. Nhita, Adiwijaya, S. Annisa and S. Kinasih, Comparative Study of Grammatical Evolution and Adaptive Neuro-Fuzzy Inference System on Rainfall Forecasting in Bandung, 2015 3rd International Conference on Information and Communication Technology (ICoICT), Nusa Dua, (2015), 6-10.

[9] Pagán, J, Risco-Martín, J.M. Moya, J and L. Ayala, J, Grammatical Evolutionary Techniques for Prompt Migraine Prediction, Proceedings of the Genetic and Evolutionary Computation Conference 2016, Denver, (2016) , 973-980.

[10] J. C. S. Hilburg, M. Zapater, J. L. R. Martín, J. M. Moya and J. L. Rodrigo, "Using Grammatical Evolution Techniques to Model the Dynamic Power Consumption of Enterprise Servers," 2015 Ninth International Conference on Complex, Intelligent, and Software Intensive Systems, Blumenau, (2015), 110-117.

[11] S Hyndman, R. J., & Athanasopoulos, G. Forecasting: principles and practice, (2016), Heathmont: OTexts.

[12] R. J. Hyndman and A. B. Koehler, “Another look at measures of forecast accuracy,” International journal of forecasting, vol. 22, no. 4, (2005). 21.

[13] Suyanto, Evolutionary Computation : Komputasi Berbasis Evolusi dan Genetika, (2013), Informatika, Bandung