International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2020): 7.803

Symbolic Regression with Genetic Programming in Forecasting Population Growth of the Philippines

Teotima Evangelista Gorres - Abato

Department of Applied Mathematics, College of Arts and Sciences, Agusandel Sur State College of Agriculture and Technology
E-mail: *teotimaabato[at]gmail.com*

Abstract: ***Human population is chaotic (complex) and the people in both systems are equally complex (dynamical system). Thus, the theory of dynamical system best explains the growth of human population. Symbolic regression (SR) with genetic programming (GP) is a model which uses the ideas of biological evolution to handle a complex problem in a dynamical system. Many prediction techniques were introduced and used by different researcher especially in the prediction of population. The Philippine Statistics Authority (PSA) used cohort model in predicting the population of the Philippines and uses birth rate and death rate in the national projection. This paper predicts the population of the Philippines using symbolic regression via genetic programming (GP) model and uses five demographic characteristics namely; birth rate, death rate, family planning methods adapted, life expectancy and fertility rate. For generating such model Eureqa software was used. The predicted value of population using the proposed population model was compared to the forecasted value from World Bank and PSA for the year 2010-2015.It was verified in the year 2015 when the PSA conducted the national census, and it was found out that the prediction value was much closer to the census result of 100,981,437 people.***

Keywords: Symbolic regression, Genetic programming, dynamical system, biological evolution, complex system

1. Introduction

Symbolic regression (SR) genetic programming (GP) is a stochastic (Poli, et. al, 2008) iteration techniques that propagate all possible symbolic models as valid (Vladislavleva, 2008) mathematical expression in a given set of input variables, basic functions and constants and searches (Lima Neto, 2009) a model (set of models) optimizing fitness objective such as prediction accuracy of the training set (observed data set).

This paper use symbolic regression (SR) with genetic programming (GP) model in predicting future values of Philippine population with the concept of dynamical system. Genetic programming is an evolutionary computation techniques that are then (modified) evolved using an evolutionary algorithm used especially in big data. Genetic programs are programs that are based on genetic algorithm that use as the tools and procedures of an evolutionary biology. In evolutionary biology, life evolve from easy to chaotic by normal choice and whose "fitness qualities - to-survive are high. They are retained and paired with other individuals of high fitness values. The offspring are generated such that they have high fitness values than their parents. The algorithm stops when an offspring is found possessing the fitness score which is pre-determined.

Langdon and Poli (2012) averred that genetic programming (GP) is an effective search algorithm especially when the target function or fitness function is not smooth. Schmidt (2013) moreover believed that genetic programming (GP) contends with standard maximization procedures when the target function is differentiable.

Goodson (2013) characterize dynamical system as the investigation of how things change with consistently shifting time. He referred to different application, for example, the development of populations, the adjustment of the climate condition, radioactive rot, blending of fluids and gases like the sea flows, movement of the planets, the premium in a financial balance, and so forth.

Gorres Evangelista (2014) proposed and evaluated mathematical model using Symbolic Regression with Genetic Programming in forecasting population growth of the Philippines in the year 2015.

In this study, we attempt to predict the Philippine population from 2010 to 2015 given the previous population values from 1960 to 2013. Each datum is assumed to be the result of the following characteristics: (a) birth rate (b) death rate (c) fertility rate (d) life expectancy (e) family planning methods adapted. We show that using genetic programming, predicted population values have lower mean-squared errors.

2. Methods and Design

The study made use of the descriptive methods of research using evolutionary algorithm. Data were obtained from the Philippine Statistical Authority Census (1960-2013). The maximization can be quantitatively measured by:

$$Minimize: MSE = \frac{\sum_{i=1}^{n}(X_i - \hat{X}_i)^2}{n-1}$$

Where: $\hat{X}_i$= predicted value of x_i using five (5) demographic characteristics.

The demographic characteristics are:
y_1= birth rate
y_2= death rate
y_3= family planning method adapted
y_4= life expectancy
y_5= fertility rate and
$\hat{X} = w_1y_1 + w_2y_2 + w_3y_3 + w_4y_4 + w_5y_5$
Where: w_1, w_2, w_3, w_4, $and w_5$ are weights to be determined by Symbolic Regression via genetic programming (GP).The program can generate the GP search process and established proposed population model. The mean absolute error (MAE) and mean squared error (MSE) are the fitness measure used

to determine the model and the desired parameter. The software used is the licensed version of EUREQA

3. Results and Discussion

From the data of population in Bangko Sentralng Pilipinas (BSP) and Philippines Statistics (PSA) from: 1960 to 2013. Each datum is assumed to be the result of the following characteristics: (a) birth rate, (b) death rate, (c) family planning methods adapted, (d) life expectancy, and (e) fertility rate. The fitness measure used in the study is the mean absolute error (MAE) and mean squared error (MSE).We show that using symbolic regression via GP, the model or mathematical formula with minimum fitness values were selected and considered as the proposed model that best fit the data. The equation (model) below is the result of SR Program or GP methodology with the given demographic characteristics as mention above.

The resulting SR via GP equation (model):

$$\hat{X} = \mathbf{9.500781328} * y_4 + \mathbf{2.155133924} * y_5 + \mathbf{0.01230745492} * y_3 + \mathbf{1.561527447} * y_1 * y_2 - \mathbf{297.578539445888} - \mathbf{7.65381673272332} * y_1 - \mathbf{46.8922778266306} * y_2 - \mathbf{0.04458296716} * y_1^2$$

Table 1: Fitness measure with different generations

Time	Number of Generations	MSE	MAE
1 second	262	0.00017315525	0.010466431
2 seconds	507	0.00017315525	0.010466431
3 seconds	594	0.00009754245	0.0062827733
4 seconds	794	0.00017315525	0.010466431
5 seconds	841	0.00017315525	0.010466431
6 seconds	922	0.00017315525	0.010466431
7 seconds	1128	0.00017315525	0.010466431
8 seconds	1435	0.00017315525	0.010466431
9 seconds	1579	0.00017315525	0.010466431
10 seconds	1849	0.00017315525	0.010466431
11 seconds	1807	0.00017315525	0.010466431
12 seconds	2135	0.00017315525	0.010466431
13 seconds	2324	0.00017315525	0.010466431
14 seconds	2357	0.00017315525	0.010466431
15 seconds	2542	0.00017315525	0.010466431

In the table 1 above the fitness measure (MAE and MSE) which are the basis for generating the GP model or computer program have similar fitness value for a different time to run the program. The model with the minimum fitness value is considered as the result of the run. It is also considered as the best model that can fit the given data. Using the above equation (model) the resulting value of the population is being validated and compared to the prediction value and census result of PSA, Bangko Sentralng Pilipinas (BSP) and World Bank as shown in the table 2 below.

Table 2: Validation and Comparison of the data of the proposed population model to the actual or census result and predicted value of Philippines population in different government agencies

Year	GP model (in millions) Predictions	World bank (in millions) Predictions	PSA Census result	PSA (in millions) Predictions	Bangko Sentralng Pilipinas Predictions (in millions)
2010	91.72061	93.727	92.34	93.135	92.6
2011	93.26012	95.278	No data	94.824	94.8
2012	94.69804	96.867	No data	96.511	97.1
2013	96.62137	98.481	No data	98.197	98.8
2014	98.39932	100.102	No data	99.880	100.5
2015	101.356	101.717	100.98	101.562	102.2

The forecasted value of the SR with GP model is very close to the prediction and census (actual) result of the government agency which is the Philippines Statistics Authority (PSA). This government agency is responsible for gathering information for the prediction and actual results of the number of people living in the Philippines. The SR-GP model with a minimum fitness value was considered as the best-fit model and used to forecast the growth of population of the Philippines in the year 2010- 2015. This implies that the SR-GP which is a machine learning based method can compete the other method in forecasting time series data particularly in a huge data with promising results. This method can also be used especially in a complex problem.

The PSA used cohort model and two demographic factors namely; birth and death rate to forecast the national population of the Philippines while the GP model used five demographic factors as enumerated above. Based from the result of the study family planning adapted as it was imposed by the government did not lessen the national population. Even family planning method was used by many Filipinos the population of the Philippines still increasing.

References

[1] Gorres Evagelista, Teotima (2014). A Genetic Programming Approach in forecasting Population Growth of the Philippines. Unpublished Master's Degree Thesis

[2] Goodson, G.R., (2013). Lecture Notes on Dynamical System, Chaos and Fractals geometry. Retrieved: January 16, 2019. fromhttp://www.issp.ac.ru/ebooks/books/open/Chaos_Fractal_Geometry.pdf

[3] Poli, R., et.al (2008). A Field Guide to Genetic Programming. Retrieved: January 16,2019 from http://www0.cs.ucl.ac.uk/staff/W.Langdon/ftp/papers/poli08_fieldguide.pdf

[4] Schmidt, Michael (2013). Latest Version of Eureqa Langdon, W. and Poli,R.,(2012).Foundation of Genetic Programming.

[5] LimaNeto, E., (2009). Beyond the non-probabilistic Symbolic regression linear regression and GP to Approach measured climatic data in a river

[6] Vladislavleva, E. (2008).*Models- based problem solving through Symbolic Regression via GP.Retrieved:*

https://www.researchgate.net/publication/46432703_Model-Based_Problem_Solving_through_Symbolic_Regression_via_Pareto_Genetic_Programming

Websites

[7] Philippine Population Would Reach Over 140 Million by the Year 2040 (Final Results from the 2000 Census-based Population Projections). http://www.census.gov.ph/statistics/census/projected-population (accessed April 4, 2006).

[8] World Bank.Crude Birth Rate of the Philippines. http://data.worldbank.org/indicator/SP.DYN.CBRT.IN (accessed April 13, 2014) http://www. PSA.gov.ph

[9] Trading Economics/BangkoSentralngPilipinas. The Philippine Population. http://www.tradingeconomics.com/philippines/population(2014).

[10] WorldBank. Contraceptive Prevalence of the Philippines. http://data.un.org/Data.aspx?d=WDI&f=Indicator_Code%3aSP.DYN.CONU.ZS(accessed February 1, 2014)

[11] Philippines.http://data.worldbank.org/indicator/SP.DYN.CBRT.IN (accessed April 13, 2014)

[12] WorldBank.Death Rate of the Philippines.http://data.worldbank.org/indicator/SP.DYN.CBRT.IN(accessed April 13, 2014)

[13] WorldBank. Total Life Expectancy of the Philippines. http://www.quandl.com/WORLDBANK/PHL_SP_DYN_LE00_IN-Philippines-Life-expectancyat-birth-total-years (accessed January 11, 2014).

Software use:

[14] Schmidt, M., and Lipson, H., (2013). Eureqa (version 0.98 beta). Available from http//www.eureqa.com