A STOCHASTIC DYNAMICAL SYSTEM MODEL OF AN EARTHQUAKE IN AGUSAN DEL SUR REGION: GENETIC PROGRAMMING BASED SYMBOLIC REGRESSION

Teotima Evangelista Gorres-Abato

Department of Applied Mathematics

Agusan del Sur State College of Agriculture and Technology,
Bunawan, Agusan del Sur, Philippines.

Abstract: The study proposes a new approach to genetic programming (GP) based on symbolic regression (SR) on a statistical search which is a topological dynamical system identification technique that uses a stochastic process for finding the earthquake models from the space (phase or state space) and the law of evolution of such functions or models in a dynamical system. It is an evolutionary algorithm aimed to recapitulate a set of measures in an n- dimensional space, used to determine the relationship of measurable functions or model (sets of functions or distributions) and parameters simultaneously unlike in traditional regression analysis only the parameters are determined while the model is defined in advanced. Since the occurrence of an earthquake cannot be predict, this study introduce an earthquake model of Agusan Del Sur.

Index Terms - Symbolic regression, stochastic process, earthquake, topological dynamical system, dynamical system, evolutionary algorithm, and genetic programming

I. INTRODUCTION

To date, the occurrence of an earthquake is unpredictable since its occurrence is randomly occurring that makes it chaotic. Many research studies were established, but still there were none could describe when the earthquake will occur. The study is about the use of genetic programming (GP) based symbolic regression (SR) attempting to make an earthquake model of a time series data. SR with GP is an evolutionary computation technique that are then (modified) evolved using an evolutionary algorithm which is used especially in a dynamical system.

Definition 1: Symbolic regression (SR) is another type of regression analysis uses stochastic process for finding the set of functions or probabilistic (mathematical) models from the space (phase or state space) of a topological dynamical system and the law of the evolution of such functions or models which uses the concept of symbolic dynamics in a dynamical system. It is an evolutionary algorithm aimed to recapitulate a set of measures in a $n-$dimensional space, use to describe the relationship of measurable functions (sets of functions) or distributions or variables for predictions of unknown samples, study the statistical performance of a dynamical system (topological dynamical system) otherwise determine the statistical properties of the system by applying the ergodic theory principle.

Given $\{x_1, x_2, \dots\dots, x_n, E\}$ where: $x_1, x_2, \dots\dots, x_n$ are random variables.

$$E = f(x_1, x_2, \dots\dots, x_n) \qquad (1)$$

The task of SR is to discover the linear or nonlinear symbolic functions or models for the unknown function, from the given sets of data in a dynamical system. SR is a stochastic process of finding a set of functions or distribution by which the model and parameters are determined simultaneously, unlike in traditional regression which seeks to optimize the parameters for a pre specified model (Gorres Evangelista (2021)).

Definition 2. Genetic Programming (GP) is a machine learning techniques that uses the concept of stochastic dynamical system to solve Symbolic Regression (SR) problem.

Defintion 3. A Dynamical Systems input – output model is an operator

$$f: M \rightarrow M$$

Let M as a "phase space" (e.g. the attractor of the system) of possible states of the system, and the map f as the "law of evolution" of the system. Then, given an "initial condition" $x_0 \in M$ have a sequence given by $x_1 = f(x), x_2 = f^2(x) = f(f(x)) = f \circ f(x)$ and generally
$x_n := f^n(x) = f \circ \dots\dots \circ f(x)$ given by the $n'th$ composition of the map f with itself (Luzzatto, S., (2018)).

Definition 4. A stochastic dynamical system is a rule for time evolution on a state space biased to the effects of noise.

Definition 5. Topological dynamical system is a qualitative, asymptotic properties of curves or model in a dynamical systems based on general topology focuses on a topological space (Dontwi, I., Obeng-Denteh,W., Neumann, F., and Opoku, S., (2011))

Goodson, (2013) defines dynamical systems as the study of how things change over time with continuously varying time (known as flows). He cited various applications

such as; the growth of populations, the change in the weather condition, radioactive decay, mixing of liquids and gases such as the ocean currents, motion of the planets, the interest in a bank account, seismic signal etc.

Langdon, and Poli (2012) averred that genetic programming (GP) is an efficient search algorithm particularly when the objective function or fitness function is not smooth (not differentiable). Schmidt (2013) likewise said that genetic programming (GP) competes with standard optimization techniques when the objective function is a differentiable function.

Asim, Alvarez, Basit, Iqbal, (2017) used machine learning in predicting earthquake magnitude in Hindu Kush and concluded that this technique had momentous and promising results. Whereas, Asima, Idrisb, Iqbala, Martínez-Álvarez, (2018) affirmed that a prediction system for an earthquake by means of combining seismic signs along with Genetic Programming (GP) and AdaBoost (GP-AdaBoost) a procedure in improving Earthquake Predictor system (EP-GPBoost) indicates greater results in terms of accuracy, precision and other Correlation performance as compared to classical results.

Kannan, Suganth,(2018) introduced an innovative mathematical model (IMM) which uses earthquake data with Latitude, Longitude and Magnitude as variables. The model uses the concept of Poisson's distribution and Spatial Connections for an enormous group of information for every quake zone. He discovered that there is a recognizable pattern inside the random events of the quakes around each fault zone.

This study attempt to introduce an earthquake model of Agusan Del Sur. Each datum is assumed to be the result of the following four (4) parameters in a dynamical system:

(a) Mean temperature (b) rainfall (c) air pressure (d) sea pressure. We show that using genetic programming based symbolic regression an earthquake model can be derived.

II. METHODS AND DESIGN

In a dynamical system, the Symbolic regression is a function that describes the relation of the input to output data and can be written as:

$$f(x_1, x_2, x_3, .., x_n) = E \qquad \rightarrow (input\ data)$$

The task of symbolic regression with Genetic programming is to identify the target expression (function) or model in symbolic form which estimates the values of a specified target variable based on the values of a set of input variables, that is;

$$\hat{E} = \hat{f}(x_n) \rightarrow output\ data$$

The study made use of the descriptive methods of research using evolutionary algorithms particularly the genetic programming based symbolic regression. Data were obtained from the Phivolcs and PAGASA Caraga. The maximization can be quantitatively measured by:

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

Where: $\hat{E}_i$= predicted value of E_i using four (4) parameters.

The parameters are:

x_1= mean temperature

x_2= rainfall

x_3= air pressure

x_3= sea pressure

$E = earthquake\ magnitude$

$$\hat{E}_i = = w_1x_1 + w_2x_2 + w_3x_3 \qquad (2)$$

Where: w_1, w_2, w_3 are weights to be determined by using Symbolic Regression via genetic programming (GP).The program can generate the GP search process and establish proposed model. The software used is the license version of DATAROBOT.

IV. RESULTS AND DISCUSSION

The analytic solution to the model using the Symbolic regression with Genetic programming mathematically expressed by:

$$\hat{E} = f(x_1, x_2, x_3)$$

The resulting model:

$$\hat{E} = 2.59 + 0.06 * cos(sqrt(x_2)) - 0.04 * cos(2.69 + x_3 - x_2 * cos(sqrt(x_2))) \qquad (3)$$

Equations (3) is the model of an earthquake in the Region of Agusan del Sur. From the four (4) parameters used in the study only two (2) parameters shown in the output model as the result of the run. In the same way, using the model as in equation (3) above, the prediction of an earthquake in some part of Agusan del Sur Region is 2.7 for the month of July 2021. Further research is recommended for the improvement of the model.

REFERENCES

[1] Asim, K.M., A´ lvarez F.M., Basit, A., Iqbal, T., (2017) . Earthquake magnitude prediction in Hindukush region using machine learning techniques. Nat Hazards (2017) 85:471–486 DOI 10.1007/s11069-016-2579-3

[2] Asima,K.M., Idrisb, A., Iqbala,T., Martínez-Álvarez, F., (2018). Seismic indicators based earthquake predictor system using Genetic Programming and AdaBoost classification. Soil Dynamics and Earthquake Engineering Journal. journal homepage: www.elsevier.com/locate/soildyn. Retrieved Jan 2019 from http://iranarze.ir/wp-content/uploads/2018/09/E9269-IranArze.pdf.

[3] Dontwi, I., Obeng-Denteh,W., Neumann, F., and Opoku, S., (2011). A study of topological dynamical systems.

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

[5] Gorres Evangelista, T., (2021). Seismic signal model in Agusan Del Sur through symbolic regression with genetic programming: A Stochastic dynamical system approach.

[6] Langdon, W. and Poli,R.,(2012).Foundation of Genetic Programming.

[7] Kannan, Suganth,(2018). Innovative Mathematical Model for Earthquake Prediction. Student, American Heritage School, Plantation, Florida, USA. Retrieved January 2019 from https://www.researchgate.net/publication/280764638_Innovative_Mathematical_Model_for_Earthquake_Prediction pl208281@ahschool.com

[8] Luzzatto, S., (2018). Introduction to Dynamical Systems. Lecture Notes. https://people.maths.bris.ac.uk/~ip13935/dyn/SLI.pdf

[9] Schmidt, Michael (2013). Latest Version of Eureqa

Software use

DataRobot . Version V f0e8f0f7a9b51308. (2021). https://app2.datarobot.com/"