Improving the Generation of Earthquake Risk Models Using Evolutionary Algorithms

Yuri Lavinas[1], Marcelo Ladeira[3], and
Claus Aranha[5]

[1] University of Brasilia, Computer Science Department   `yclavinas@gmail.com`
[2] University of Brasilia, Computer Science Department   `mladeira@unb.br`
[3] University of Tsukuba, Graduate School of SIE   `caranha@cs.tsukuba.ac.jp`

Abstract. Earthquake Risk Models describe the risk of occurrence of seismic events on an area based on information such as past earthquakes in nearby regions and the seismic properties of the area under study. These models can be used to help to better understand earthquakes, their patterns and their mechanisms. In previous work, we showed that Genetic Algorithms can be used to generate earthquake risk models. In this study we want to analyze any similarities between generating earthquake risk models and the between exploring the search space of BBOB benchmark functions both with Genetic Algorithms. This is done by examining the results diferent values for the selection operator through simulations using the catalog of Japanese earthquakes and the 24 BBOB noise free benchmarks functions with 40 dimensions and comparing their behavior. The results are XXX (here is where i should state that the tournsize = 2).

1   Introduction

Earthquakes can cause great damage to human society through soil rupture, movement, tsunami, etc. Some recent earthquakes that highlight this destructive potential are the great East Japan Earthquake of 2011 (Figure 1), and the April 2015 earthquake in Nepal. One important tool for the enactment of policies that minimize the consequences of these events are earthquake occurrence models (also called risk models). These models can be used to identify patterns in the seismic mechanisms that generate earthquakes, and are important to increase our understanding of these events.

In our previous work [1], we proposed a way to generate earthquake risk models using a standard Genetic Algorithm (here called the GAModel). The GAModel was shown to be competitive with the Relative Intensity (RI) model. In this paper, the main goal is to observe the performance accomplished by the GAModel and a real-valued Genetic Algorithm (based on the GAModel, GA-BBOB) applied to 24 noise-free 40 dimensions BBOB benchmark functions. To achieve that, we first explore the number of individuals to be selected by the Tournament operator. Then we analyze the results of the GAModel and the results of the GA-BBOB to verify any relationship between applying the
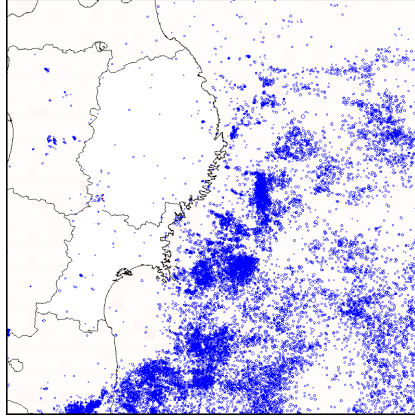
Fig. 1: Seismic Activity in Eastern Japan in 2011. Each dot shows one earthquake

Tournament operator in the "generating earthquake risk model" function and applying tht Tournament operator in all the benchmark functions.

These adaptions are described here and there. we do this and that here and there too. results and other stuff are here and there.

2   Background

An Earthquake Risk model states the probability of earthquake occurrence on a defined area and time period. These models are often based on past occurrence of earthquakes (historical catalogs). They can also make use of seismic properties of the area under study, such as faults and terrain composition.

The "prediction" of earthquakes is a polemic subject, and no research so far has come close to suggesting that individual large scale earthquakes can be predicted. On the other hand, there is value on the study of earthquake mechanisms and the generation of statistical models of earthquake risk [14].

In our previous work [1], we use a Genetic Algorithm (GA) to optimise an Earthquake Risk Model, following the framework proposed by the Collaboratory for the Study of Earthquake Predictability (CSEP).

The CSEP framework defines a model in reference to a geographical region and a time period [18]. The geographical region is divided in a grid, where each cell in the grid is called a bin.

The model defines a number of expected earthquakes for each bin. This number must be a positive integer. A good model is one where the number of estimated earthquakes in each bin corresponds to the actual number of earthquakes that occurs in that bin during the target time interval.

2.1 The GAModel

Using the CSEP framework described in the previous subsection, we proposed the GAModel [1], which uses Genetic Algorithms to generate an earthquake risk model based on earthquake catalog data.

In the GAModel, each individual is treated as a prediction in the CSEP framework. The fitness of each individual will be calculated using the log-likelihood of the catalog data given the individual's prediction.

Genome Representation and Evolutionary Operators. Each individual is represented as real valued array, where each element is a bin, with an associated number of earthquakes. One-point crossover, elitism and polynomial bounded mutation are used as evolutionary operators. The relevant parameters were set as Elite Size = 1, Crossover chance = 0.9, Mutation Chance = 0.1, Polynomial Bounded parameters eta = 1, low = 0, up = 1.

Fitness Function. Let an individual $\Lambda = \{\lambda_1, \lambda_2, \ldots, \lambda_N\}$ be a forecast in the CSEP framework. Let the set of earthquake occurrences from the catalog be $\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\}$. The log-likelihood of the catalog data given an individual is calculated as:

$$L(\Omega|\Lambda) = \sum_{i=1}^{n} L(\omega_i|\lambda_i) = \sum_{i=1}^{n} -\lambda_i + \omega_i \log \lambda_i - \log \omega_i!. \tag{1}$$

To avoid overfitting, the period under consideration is divided into sub-periods, the log-likelihood for each sub-period is calculated separately, and the worst value is used as the fitness [1].

2.2 Related Literature

The usage of Evolutionary Computation (EC) in the field of earthquake risk models is somewhat sporadic. Zhang and Wang [19] used Genetic Algorithms to fine tune an Artificial Neural Network (ANN) and used this system to produce a forecast model. Zhou and Zu [20] also proposed a combination of ANN and EC, but their system forecasts only the magnitude parameter of earthquakes. Sadat, in [13], used ANN and GA to predict the magnitude of the earthquakes in North Iran.

There are more works when we discuss EC methods and estimation of parameter values in seismological models. Nicknam et al. [10] simulated some components from a seismogram station and predicted seismograms for other stations. They combined the Empirical Green's Function (EGF) with GA. Kennett and Sambridge [8] used GA and associated teleseisms procedures to determine the Fault Model parameters of an earthquake.

Another popular approach is to use EC methods do calculate the Peak Ground Acceleration (PGA) parameter. The works done by Kerh et al. [9, ?]

are a combination of ANN and GA to estimate or predict PGA in Taiwan. Cabalar and Cevik [2] work also aimed to predict the PGA, but their work uses genetic programming (GP) and use strong-ground-motion data from Turkey.

Jafarian et al. [7], used GP to develop an empirical predictive equation v_max/a_max ratio of the shallow crustal strong ground motions recorded at free field sites. They found a relation between the v_max/a_max and the earthquake magnitude and the source-to-site distance.

Ramos and Vázques [12] used Genetic Algorithms to decide the location of sensing stations. In this work they achieved, in general, better results with the GA method when compared with the Seismic Alert System (SAS) method and a greedy algorithm method.

Saeidian et al. [15] work also based on the same idea of locating sensing stations. They do a comparison in performance between the GA and Bees Algorithm (BA) to decide which of those techniques would perform better when choosing the location of sensing stations. He found out that the GA was faster than the BA.

Huda and Santosa [6] published a paper in which the goal was to find, via GA, the speed of the waves P and S in the mantle and in the earth crust. P waves are indicated as the first fault found in seismological data and S waves are the changes caused in the phase of a P wave [6]. This work aimed to obtain a structure of the Japanese underground.

3   GA-BBOB

In the work, we propose the GA-BBOB, which is a real-valued Genetic Algorithm to explore the search space of the 24 noise free 40-D BBOB benchmark functions. In some aspects the GA-BBOB implementation is equal to the GAModel. That is mainly true for all evolutionary operators. The search space of the benchmark functions are distinct from the GAModel search space, we reformulated the Genome Representation and the Fitness Function.

Genome Representation and Evolutionary Tools. Each individual is represented as real valued array, where each element is one input to a 40-D noise free benchmark BBOB function. Therefore, each individual of the GA-BBOB has size of 40 real value elements. Each element's values has values between the interval [-5, 5], once those function are limited by these values. As in the GAModel it is used the Two-Point crossover, elitism and polynomial bounded mutation are used as evolutionary operators. The relevant parameters were set as Elite Size = 1, Crossover chance = 0.9, Mutation Chance = 0.1. For the Polynomial Bounded parameters, a small modification, when comparing to the GAModel, was made: eta = 1, low = -5, up = 5.

Fitness Function. The fitness function considered are the 40-D noise free benchmark BBOB function. For more information, please refer to the work of Hansen et al. [4]. Consequently, the GA-BBOB has 24 different fitness functions.

Minor Adaptations. Chuang et al. [3], proposed a Genetic Algorithm specific to run on the noise free BBOB benchmark functions. Some good practices presented in this work were incorporated to the GAModel and to the GA-BBOB. These practice are the concept of "restart strategy" and the concept of "alleviate stagnation".

The restart strategy concept is defined by Chuang et al. as: "For each restart, the initial population is uniformly and randomly sampled within the search space. Whenever the restart condition is met, the algorithm will be reinitialized without using any information about the last test run. This process is iterated until the stopping criteria are met, i.e., maximum number of function evaluations has been reached, or the function value is less than the target precision."

The restart condition used for both GAModel and GA-BBOB is that the standard deviation of the current population is smaller that 10^{-12}. When the restart condition is contemplated the algorithm population is reinitialized randomly, as in the first initialization. In the GAModel and in the GA-BBOB, we consider information about the last run, by utilizing the Elitism strategy on the best individual prior to the restart. The maximum number of function evaluations chosen is (10^5) and the target precision chosen is (10^{-8}), values compatible with the ones in the previous work.

3.1 Genetic Algorithms and Benchmarks Functions

IN 2009, Nicolau [11] proposed the first Genetic Algorithm on the BBOB-2009 noiseless testbed. It was a simple binary Genetic Algorithm and its results showed good results on separable functions, but poor performance is achieved on the other functions.

In 2012, Chuang et al. [3] presented the "DBRCGA", a real coded Genetic Algorithm that uses relative fitness information to direct the crossover toward a direction that significantly improves the objective fitness. The DBRCGA was tested on the BBOB-2012 noiseless testbed. Their results showed that the DBRCGA performs with difficulty in getting a solution with the desired accuracy for high conditioning and multimodal functions within the specified maximum number of evaluations, the DBRCGA presents good performance in separable function and functions with low or moderate conditioning.

In 2013, Holtschulte et al. [5] evaluated two Genetic Algorithm on the BBOB-2013 noiseless testbed, but none had results competitive with the best 2009 optimization algorithm. Their results highlight the importance of carefully chosen genetic operators.

Sawyerr et al found simliar results to the one achieved by Chuang et al. on their works in 2013 [17] and in 2015 [16]. In 2013, they proposed the "PRCGA", or projection-based real-coded genetic algorithm. It incorporates exploratory search mecanism based on vector projection. In 2015, they have studied a hybrid application of the Genetic Algorithm (GA) with the *uniform random direction* search, named "RCGAu". BOth the PRCGA and the RCGAu were tested on the BBOB-2013 noiseless testbed. They stated that both of them had difficulties to achieve solutions with the desired accuracy for high conditioning and multimodal

functions, though they were able to solve the benchmark functions. Sawyerr et al concluded that the RCGAu has excelled in solving the f_1, f_2, f_3, f_7 and f_{21}, tough for the other functions it achieved avarage results. They also state that real value GA do not efficiently solve highly conditioned problems and studies have currently been carried out to find out why [16].

4 Experiment Design

4.1 Experiment Design

4.2 Data Sets - GAModel

We use the earthquake catalog made available by the *Japan Meteorological Agency* (JMA) webpage. From this catalog, we use the following earthquake data: time of occurrence, latitude, longitude and we focus on one area for our study, the "East Japan" area.

East Japan region covers the east coast of Japan, including a large ocean area. In this work we define it as the area within latitude 37N to 41N, and 140W to 144W. It is divided into 1600 bins of approximately 100km^2.

5 Results

6 Discussion

7 Conclusion

References

1. Aranha, C., Lavinas, Y.C., Ladeira, M., Enescu, B.: Is it possible to generate good earthquake risk models using genetic algorithms? In: Proceedings of the International Conference on Evolutionary Computation Theory and Applications. pp. 49–58 (2014)
2. Cabalar, A.F., Cevik, A.: Genetic programming-based attenuation relationship: An application of recent earthquakes in turkey. Computers and Geosciences 35, 1884–1896 (October 2009)
3. Chuang, Y.C., Chen, C.T.: Black-box optimization benchmarking for noiseless function testbed using a direction-based rcga. In: Proceedings of the 14th annual conference companion on Genetic and evolutionary computation. pp. 167–174. ACM (2012)
4. Hansen, N., Auger, A., Finck, S., Ros, R.: Real-parameter black-box optimization benchmarking 2010: Experimental setup. Ph.D. thesis, INRIA (2010)

5. Holtschulte, N.J., Moses, M.: Benchmarking cellular genetic algorithms on the bbob noiseless testbed. In: Proceedings of the 15th annual conference companion on Genetic and evolutionary computation. pp. 1201–1208. ACM (2013)

6. Huda, A.M., Santosa, B.: Subsurface structure in japan based on p and s waves travel time analysis using genetic algorithm in japan seismological network. International Journal of Science and Engineering 6(1) (2014), `http://www.ejournal.undip.ac.id/index.php/ijse/article/view/5762`

7. Jafarian, Y., Kermani, E., Baziar, M.H.: Empirical predictive model for the v max/a max ratio of strong ground motions using genetic programming. Computers & Geosciences 36(12), 1523–1531 (2010)

8. Kennet, B.L.N., Sambridge, M.S.: Earthquake location — genetic algorithms for teleseisms. Physics of the Earth and Planetary Interiors 75(1–3), 103–110 (December 1992)

9. Kerh, T., Gunaratnam, D., Chan, Y.: Neural computing with genetic algorithm in evaluating potentially hazardous metropolitan areas result from earthquake. Neural Comput. Appl. 19(4), 521–529 (Jun 2010), `http://dx.doi.org/10.1007/s00521-009-0301-z`

10. Nicknam, A., Abbasnia, R., Eslamian, Y., Bozorgnasab, M., Mosabbeb, E.A.: Source parameters estimation of 2003 bam earthquake mw 6.5 using empirical green's function method, based on an evolutionary approach. J. Earth Syst. Sci. 119(3), 383–396 (June 2010)

11. Nicolau, M.: Application of a simple binary genetic algorithm to a noiseless testbed benchmark. In: Proceedings of the 11th Annual Conference Companion on Genetic and Evolutionary Computation Conference: Late Breaking Papers. pp. 2473–2478. ACM (2009)

12. Ramos, J.I.E., Vázques, R.A.: Locating seismic-sense stations through genetic algorithms. In: Proceedings of the GECCO'11. pp. 941–948. ACM, Dublin, Ireland (July 2011)

13. Sadat, N., Zakeri, S., Pashazadeh, S.: Application of neural network based on genetic algorithm in predicting magnitude of earthquake in north tabriz fault (nw iran). Current Science (00113891) 109(9) (2015)

14. Saegusa, A.: Japan tries to understand quakes, not predict them. Nature 397, 284 (1999)

15. Saeidian, B., Mesgari, M.S., Ghodousi, M.: Evaluation and comparison of genetic algorithm and bees algorithm for location-allocation of earthquake relief centers. International Journal of Disaster Risk Reduction (2016)

16. Sawyerr, B.A., Adewumi, A.O., Ali, M.M.: Benchmarking rcgau on the noiseless bbob testbed. The Scientific World Journal 2015 (2015)

17. Sawyerr, B.A., Adewumi, A.O., Ali, M.M.: Benchmarking projection-based real coded genetic algorithm on bbob-2013 noiseless function testbed. In: Proceedings of the 15th annual conference companion on Genetic and evolutionary computation. pp. 1193–1200. ACM (2013)

18. Zechar, J.D.: Evaluating earthquake predictions and earthquake forecasts: A guide for students and new researchers. Community Online Resource for Statistical Seismicity Analysis pp. 1–26 (2010)

19. Zhang, Q., Wang, C.: Using genetic algorithms to optimize artificial neural network: a case study on earthquake prediction. In: Second International Conference on Genetic and Evolutionary Computing. pp. 128–131. IEEE (2012)

20. Zhou, F., Zhu, X.: Earthquake prediction based on lm-bp neural network. In: Liu, X., Ye, Y. (eds.) Proceedings of the 9th International Symposium on Linear Drives

for Industry Applications, Volume 1, Lecture Notes in Electrical Engineering, vol. 270, pp. 13–20. Springer Berlin Heidelberg (2014), `http://dx.doi.org/10.1007/978-3-642-40618-8\_2`