Is it possible to generate good Earthquake Risk Models using Genetic Algorithms?

Blind Author 1
University 1
Address 1
Address 2
e-mail@address.com

Blind Author 2
University 1
Address 1
Address 2
e-mail@address.com

Blind Author 3
University 1
Address 1
Address 2
e-mail@address.com

Blind Author 4
University 1
Address 1
Address 2
e-mail@address.com

ABSTRACT

Understanding the mechanisms and patterns of earthquake occurrence is of crucial importance for assessing and mitigating the seismic risk. In this work we analyze the viability of using Evolutionary Computation (EC) as a means of generating models for the occurrence of earthquakes. Our proposal is made in the context of the "Collaboratory for the Study of Earthquake Predictability" (CSEP), an international effort to standardize the study and testing of earthquake forecasting models.

We use a standard Genetic Algorithm (GA) with real valued genome, where each allele corresponds to a bin in the forecast model. The design of an appropriate fitness function is the main challenge for this task, and we test three different proposals, all based on the log-likelihood of the candidate model against the training data set.

The resulting forecasting models are compared with statistical models traditionally employed by the CSEP community, using data from the Japan Meteorological Agency (JMA) earthquake catalog. Our results indicate promise for the use of GA as basis for constructing statistical earthquake forecast models. Based on these results, we identify research directions that we consider deserve more attention from the EC community.

Categories and Subject Descriptors

I.2.8 [**Artificial Intelligence**]: Problem Solving, Control Methods and Search Heuristic Methods; J.2 [**Computer Application**]: Physical Sciences and Engineering—*Earth and Atmospheric Sciences*

GECCO'14 July 12-16,2014, Vancouver, BC, Canada.

General Terms

Real World Applications

Keywords

Earthquakes, Risk Models, Genetic Algorithms, Forecasting

1. INTRODUCTION

(Goals of the introduction, 1: Establish significance of Earthquake studies, and define the "Forecast Model" problem. Define the CESP and RELM initiatives. 2: Explain the difficulty of the Forecast Model problem and explain the motivation for use of Genetic Algorithms. 3: This is a little explored problem in GA, establish the scope of this paper. 4: Summarize the contribution and results of this paper)

** Studying earthquakes is important to society. Earthquakes are natural events that can cause large amounts of damage both in terms of loss of human life, and material damages. By understanding and measuring the risk of earthquakes, it would be possible to guide policy decisions regarding construction regulations, and the deployment of disaster relief infrastructure.

** However, the specifics of earthquake generation mechanisms are not yet completely understood. In this sense, the development of accurate models for the risk assessment of earthquake activity is a very active area of study.

** In the study of Earthquake predictability, we highlight the RELM (Relative Earthquake Log-likelihood Model) and the CSEP

**** Evolutionary algorithms have shown themselves to be very effective in hard problems, such as (...). However, the use of EC techniques in seismology studies, and in particular in the development of earthquake risk models, is extremely scarse.

** We believe that a real contribution can be make by the use of evolutionary techniques in this field. Therefore in this paper we **** make a tutorial on earthquakes,

*** We illustrate these ideas by designing a simple genetic algorithm that creates an earthquake prediction model as defined by the CSEP. We compare this algorithm with a "skilless" model (random model), and the RI model. Using Data from the JMA catalog between 2002 and 2010

** Outline of the paper In section 2, we describe the Earthquake predictability problem.

2. THE EARTHQUAKE PREDICTABILITY PROBLEM

Earthquake Risk Analysis is a huge area, which includes the development of models for earthquakes, and many other fun things. We are concentrating on the development of Earthquake Predictability Model.

In this problem, we want to create a model that, for a certain area during a certain time in the future, defines expected number of earthquakes and their respective power.

- This area is huge, we are focusing in a small sub area, which is nonetheless very intersting.

There are two main types of models: In the statistical models, the forecast is based mainly on the distribution of earthquakes of different magnitudes in the past. In Geophisical models, a theoretical model of how the seismic energy is released, and forecasts are made based on these models.

2.1 RELM And CSEP

Description of the CSEP - collaboratory for the study of earthquake predictability. It stabilish ways to analyse and compare earthquake predictability experiments.

The CSEP define a series of standardized tests, such as RELM based tests (R Test, N Test, ETC) [8], and also alarm based tests, such as the Molchamm Diagram, the Area Skill Score [9], etc.

The CSEP determines "testing windows" – one day, one month, three months, one year, three years.

2.2 Earthquake Predictability Model Testing

Model Definition – bins with a number of earthquakes in them.

Log Likelyhood comparison - Explanation

CSEP tests - Explanation

2.3 The RI algorithm

Outline of the RI Algorithm [4], and how it is often used as a benchmark

3. EVOLUTIONARY COMPUTATION FOR EARTHQUAKE RISK ANALYSIS

Compared to other application fields, there are relatively few applications of Evolutionary Computation to seismology problems. We identify

In the field of seismology and earthquake risk analysis, the few cases of Evolutionary Algorithm approaches have usually taken one of two forms. EC is often used to estimate parameter values for seismological models. These models can be used to describe and understand earthquakes.

For example, Ramos [7] uses Genetic Algorithms to decide the location of sensing stations in a seismically active area in Mexico. Nicknam et. al [5] and Kennett and Sambridge [2] used evolutionary computation to determine the Fault Model parameters (such as epicenter location, strike, dip, etc) of a given earthquake. Evolutionary Computation has also been used to estimate the peak ground accelleration of seismically active areas [3, 1].

EC has also been used very few times as a method to generate seismic forecast model. One such approach is described by Zhang and Wang [10], where they fine tune a Neural Network with a GA, and then use this system to produce a forecast. Unfortunately that work did not provide enough information to reproduce the proposed GA+ANN system or their results.

4. A FORECAST MODEL USING GENETIC ALGORITHMS

To investigate the ability of Evolutionary Computation to generate earthquake forecast models, we design and test a simple Genetic Algorithm. Let us call this system the *GAModel*.

An individual in GAModel will encode a forecast model as defined in the CSEP framework (**??**). The population will be trained on earthquake occurrence data for a fixed training period. The best generated individual will be compared with a model generated by the RI algorithm, and a skilless (random) model. This comparison is based on earthquake occurrence data for a test period immediately posterior to the training period.

By encoding the entire forecast model as one individual, we identified two main concerns while designing GAModel: First, as the forecast model contains a large number of bins, the genome of an individual will be respectively large. Evolutionary operators and parameters must be chosen carefully to guarantee convergence in a reasonable time, and avoid local optima.

Secondly, the design of the fitness function deserves a lot of attention, to avoid the risk of the system overfitting to the training data.

4.1 Genome Representation

In GAModel, each individual represents an entire forecast model. The genome is a real valued array X, where each element corresponds to one bin in the desired model (the number of bins is defined by the problem). Each element $x_i \in X$ can take a value from $[0, 1)$. In the initial population, these values are sampled from a uniform distribution.

In the CSEP framework, a model is defined as a set of integer valued expectations, corresponding to the number of predicted earthquakes for each bin. To convert from the real valued chromossome to a integer forecast, we use a modification of the Poisson deviates extraction algorithm from [6] (Chapter 7.3.12).

Algorithm 1 Obtain a poisson deviate from a $[0, 1)$ value

```
Parameters 0 ≤ x < 1, μ ≥ 0
L ← exp(−μ), k ← 0, prob ← 1
repeat
   increment k
   prob ← prob * x
until prob > L
return k
```

In Algorithm 1, x is the value taken from the chromossome, and μ is the average number of earthquakes observed across the entire training data. Note that in the original algorithm, $k - 1$ is returned. Because the log likelihood calculation used for model comparison discards models that estimate 0 events in bins where earthquakes are observed, we modify the original algorithm here to make sure all bins estimate at least one event.

4.2 Fitness Function Choice

4.3 Evolutionary Operators and Parameters

GAModel uses a regular generational genetic algorithm. For selection, we use Elitism and Tournament selection.

For the crossover operator we use *aNDX*, recently proposed by Someya [**?**]. If the value of a chromossome after the crossover falls outside the $(0, 1]$ boundary, it is truncated to these limits. For the mutation operator, we sample entirely new values from $(0, 1]$ for each mutated chromossome.

The parameters used for the evolutionary computation are as follows:

Table 1: Parameters for GAModel

Population Size	1000
Generation Number	500
Elite Size	1
Tournament Size	50
Crossover Chance	0.9
aNDX parents	4
aNDX zeta	0.5
aNDX sigma	1
Mutation Chance (individual)	0.8
Mutation Chance (chromossome)	$(\text{genome size})^{-1}$

5. EXPERIMENTS

5.1 Experimental Data

We took the data from the JMA catalog, and we divided it into regions and dates, using the following filters.

We use the following regions - All japan, an assortment of land regions, and a mixed land-sear region to see how the algorithm behaves.

Land-locked region, blablabla.

After dividing the data spatially, we divide it temporally too.

5.2 Experimental Design

*** What do we want to find out? - Is it possible to find a good model using our GA proposal? (evolution testing) - Is our simple proposal acceptable as a Predictability model? (Comparison with traditional models) *** Description of the experiments and analysis made - Data description - Time period description - Comparison metric descriptions and reasoning - Log Likelihood - Visual Comparison - AIC/BIC - Molcham Diagram - Comparison Methods (RI, Random) description and reasoning

5.2.1 Data

5.3 Results

6. CONCLUSION

– Conclusions and Future Work - Development of the current idea - Involvement of the community (De-clustering, Parameter optimization) - Please let's work with earthquakes! - GP

6.1 Future Work

7. REFERENCES

[1] A. F. Cabalar and A. Cevik. Genetic programming-based attenuation relationship: An application of recent earthquakes in turkey. *Computers and Geosciences*, 35:1884–1896, October 2009.

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

[3] E. Kermani, Y. Jafarian, and M. H. Baziar. New predictive models for the v_{max}/a_{max} ratio of strong ground motions using genetic programming. *International Journal of Civil Engineering*, 7(4):236–247, December 2009.

[4] K. Z. Nanjo. Earthquake forecasts for the csep japan experiment based on the ri algorithm. *Earth Planets Space*, 63:261–274, 2011.

[5] A. Nicknam, R. Abbasnia, Y. Eslamian, M. Bozorgnasab, and E. A. Mosabbeb. 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.

[6] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes, The Art of Scientific Computing.* Cambridge University Press, third edition, 2007.

[7] J. I. E. Rmaos and R. A. Vázques. Locating seismic-sense stations through genetic algorithms. In *Proceedings of the GECCO'11*, pages 941–948, Dublin, Ireland, July 2011. ACM.

[8] D. Schorlemmer, M. Gerstenberger, S. Wiemer, D. Jackson, and D. A. Rhoades. Earthquake likelihood model testing. *Seismological Research Letters*, 78(1):17–29, 2007.

[9] J. D. Zechar and T. H. Jordan. The area skill score statistic for evaluating earthquake predictability experiments. *Pure and Applied Geophysics*, 167(8–9):893–906, August 2010.

[10] Q. Zhang and C. Wang. Using genetic algorithms to optimize artificial neural network: a case study on earthquake prediction. In *Second International Conference on Genetic and Evolutionary Computing*, pages 128–131. IEEE, 2012.