**CSE 6242 : Data and Visual Analytics**

**Project Progress Report : Earthquake Hazard Evaluation & Visualization**

Cheng Zhang, Longchao Jia, Qiwei Mao, Tianhui Zhao, Zhen Zhong

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

Throughout the ages, as one of the major natural disasters of the world, the disaster caused by each devastating earthquake has left unforgettable memories.

From a physics point of view, the generation of natural earthquakes is caused by the breakdown of underground rocks [1]. According to the statistics [2], 50.9% of all people who died from natural disasters died from the earthquake which is highly likely to be transformed into other forms of natural disasters, like tsunami or mudslide, which can also kill many people [3].

Intuitively, people have tried to document the patterns of earthquakes since ancient times, the ultimate goal of which is to predict when, where and the intensity of earthquakes. As is known to all, it is still impossible to make such predictions, however, the risk of potential earthquake hazard of certain area can be quantified.

Among the existing earthquake prediction models, none of them can closely relate structural damage to earthquake intensity, since they build the model by using the ground motion record directly. We have extracted important parameters from the original records and combined with original data to predict parameters that are closely related to structural damage, namely the Cumulative Absolute Velocity (CAV).

For our project, specifically, we aim at developing a model that would predict the seismic damage level of a certain building based on historical and geological information gathered from free accessed databases, such as largest historical or probable earthquake magnitude (M), fault type, distance (D), and soil condition (Vs30). However, we won’t directly predict damage level, but rather an intensity measure which captures the power of the earthquakes [15]. Then, this intensity measure will be related to damage level based on ASCE standards.

Firstly, we have already managed a series of historical acceleration records of the ground during the earthquake and selecting parameters that best describe the patterns of these earthquakes [16]. For example, peak ground acceleration (PGA), Arias Intensity (IA) and CAV. It is more convenient to relate these parameters with the damage of a building rather than using pure random earthquake waves [17].

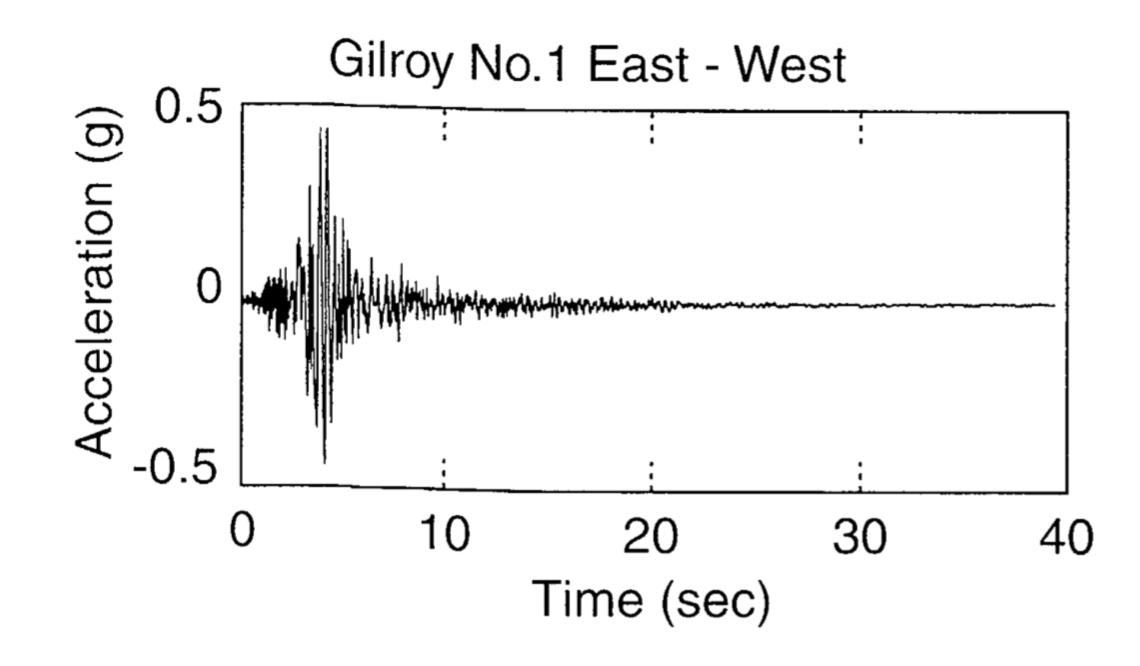


Figure 1. A typical ground motion plot, *Kramer. 1996*

This plot captures the acceleration of the ground in a certain earthquake. The largest acceleration during the event is referred to as PGA.

Secondly, for the prediction model, our problem was regarded as a regression problem with high-dimensional input, which could solve the overfitting problem. Lots of researches have been done to deal with this issue, the most obvious method is by applying the lasso or ridge regression directly [4, 9, 10, 11], or the adaptive lasso [7, 8]. Specifically, we have tried Decision Tree, Neural Network, Ada-Boosting, Support Vector Machine (SVM) and K Nearest Neighbors (KNNs) unsupervised machine learning algorithms on our current dataset. Also, we have tried to project the previous variable space to the new space using the PC regression and partial least square method, on which we shrank the dimension of the variables and reduced the dimensions[5, 6]. Moreover, after the variable selection, we have built a multiple regression model to predict the acceleration. With the comparison of the machine learning algorithms, we have decided to apply neural network and multiple regression as training methods. Once we have selected the model, we can proceed with the model evaluation and experiments.

To evaluate our model, we applied k-fold cross-validation to reduce the prediction error and prevent overfitting. After doing the features selection and applying machine learning algorithm, the data should be considered clean and organized. We will first do the test on split the dataset into k different folders to be tested separately. In order to be precise, we will test on k-folds with different k-value. After separating with k-folds, each folds data will be divided into training set and test set. Different test set proportion (from 0.1 to 0.9) will also be tested to obtain the best output fit. Once the cross-validation result is obtained, we can compare the correctness of the model and evaluate the result. As in the example figure, we will compare the minimum, maximum, average and SSD on the overall folds output for the machine learning algorithm. With the help of cross-validation, we can be more confident and precise about our prediction.

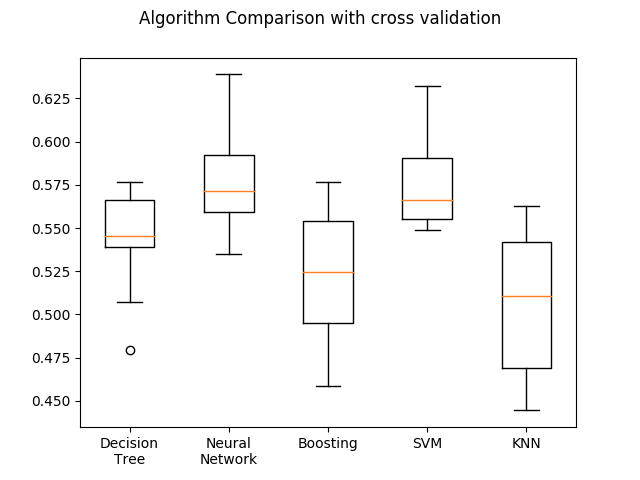


Figure 2. Example of k-folds cross-validation comparasion with k = 10

With the previous mentioned model, we can apply the predicted acceleration to a particular building and determine the damage level. In a realistic setting, however, the building could be affected by several different earthquakes, which means more than one earthquake intensity measure can be predicted by our model, which one should be used to characterize the response of our building? To address this problem, Probabilistic Seismic Hazard Analysis (PSHA) is implemented to combine all of the possible earthquakes and yield one parameter than captures all impacts. The major steps are shown in the following figure.

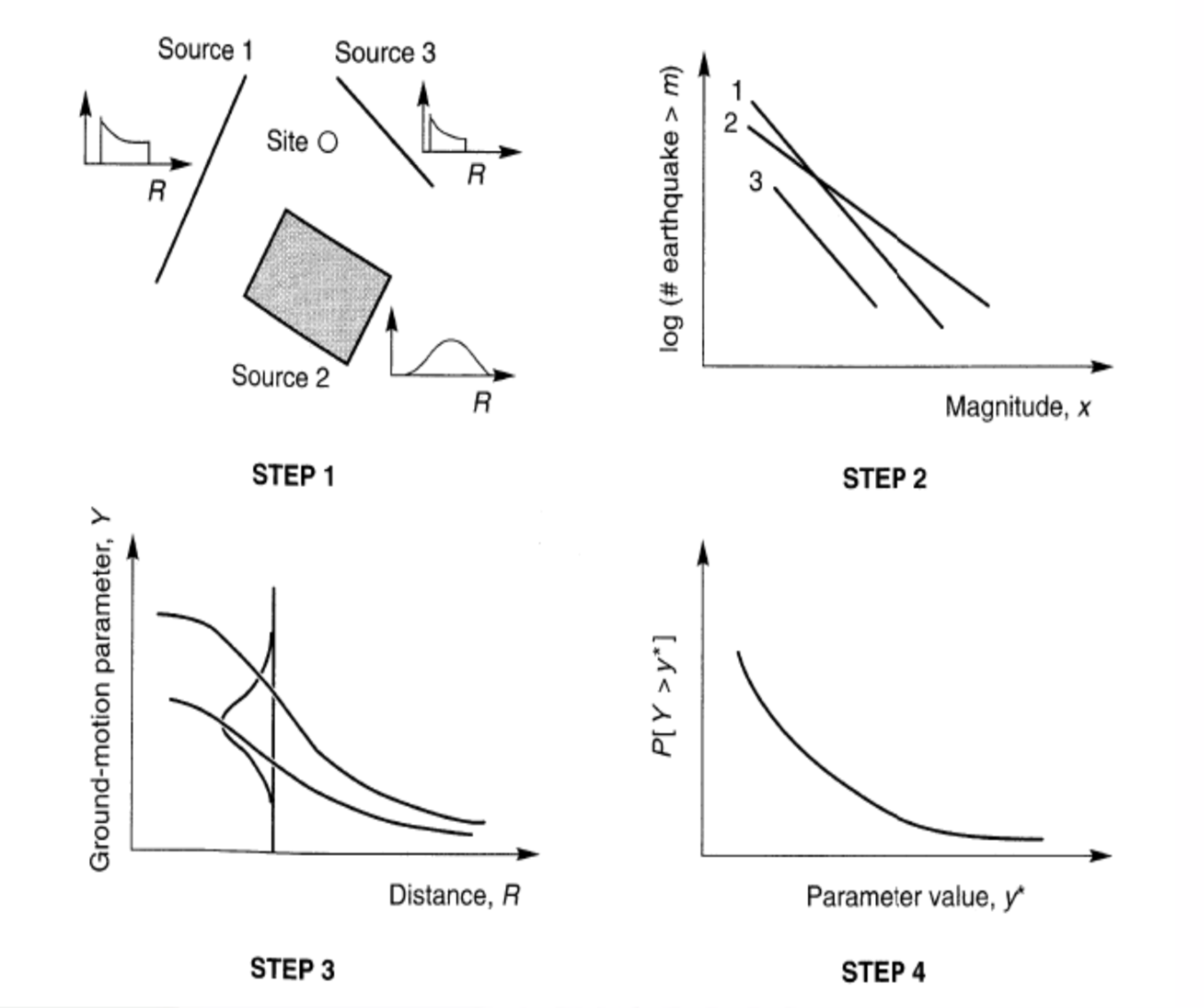


Figure 3. Main Procedure For PSHA

PSHA is a method that assumes the occurrence of earthquake follows exponential distribution, for different earthquake sources the rate of occurrence is different. When performing a hazard analysis, the exceedance of certain event within certain time period will be evaluated, for instance, “10% exceedance in 1000 years”. This analysis assumes that the occurrence of earthquake in certain period follows a poisson process. In general, PSHA can be expressed in this fashion:

Consequently, the prediction has a level of confidence, which is associated with the probability density function (pdf) of the geological fault in terms of its activity, variability and uncertainty. This value will be evaluated in every site-related analysis on our website.

In terms of damage level, there are mainly 3 degrees of damage: collapsed, still available and little to no danger [13]. In our project, we consider the building to be a 1 to 100 story reinforced concrete frame structure and compare the inter-story force, which can be calculated through SDOF analysis, on each floor with the shear strength to determine the damage level, typical shear strength can be obtained from literatures [12, 14].

For the visualization part, we are planning to display the earthquake intensity, damage level and prediction confidence data on google map.

Specifically, the visualization consists of mainly two web-pages. One is a general map that viewers can locate their mouse on a particular site on the map and click, the program will analysis the longitude and latitude of the site according to the position of the cursor, then access the corresponding data in the United States Geological Survey (USGS) and put them into the equations gotten from the machine learning model [21, 23]. The results shown on this page include the predicted intensity of the earthquake that will happen in the future, the confidence of this prediction and the recommended maximum layer of the building in this area.

The other web-page is a heatmap. Rather than the normal map on the first page, heatmaps use color and shape to represent the distribution of the data, making it easier for viewers to understand the distribution of the predicted intensity of earthquakes [22]. In the heatmap, red represents a higher intensity while green represents a lower intensity. The heatmap will also be generated based on the data from USGS and the prediction equation from the machine learning model. We intended to use the same value of the intensity in each 10\*10km2 area and smooth the color to make the heatmap beautiful, reducing computational workload in the same time.

Once our project finished, it can be used by earthquake associations as the reference of revising seismic design code, thereby insuring the earthquake resistant behavior of newly constructed buildings in a more economical way. Moreover, the program can predict the damage of current buildings, thus helping make the disaster rescue scheme wisely.

The provisional work chart and time schedule for the project is shown on the figure below. Different color represents the responsibility of each. And for our project, all team members have contributed similar amount of effort.

| **Task** | **Start Time** | **End Time** | **Team Member** | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Cheng Zhang** | **Longchao Jia** | **Qiwei Mao** | **Tianhui Zhao** | **Zhen Zhong** |
| **Surveying** | 2/7/2019 | 4/20/2019 |  |  |  |  |  |
| **Project Management** | 2/7/2019 | 4/20/2019 |  |  |  |  |  |
| **Collecting and Construct Database** | 2/15/2019 | 3/1/2019 |  |  |  |  |  |
| **Construct Prediction Model**  **(ML)** | 3/1/2019 | 3/20/2019 |  |  |  |  |  |
| **Damage Level Prediction** | 3/20/2019 | 3/30/2019 |  |  |  |  |  |
| **Data Visualization** | 3/30/2019 | 4/15/2019 |  |  |  |  |  |
| **Testing** | 4/15/2019 | 4/20/2019 |  |  |  |  |  |
| **Documentation** | 4/20/2019 | 4/23/2019 |  |  |  |  |  |

Figure 4. Work Chart & Time Schedule

**REFERENCES**

[1] N. M. Beeler, D. A. Why earthquakes correlate weakly with the solid Earth tides: Effects of periodic stress on the rate and probability of earthquake occurrence[J]. Journal of JGR Solid Earth, 2003, 108(B8).

[2] John Kelleher, John Savino, Hugh Rowlett, William McCann Why and Where Great Thrust Earthquakes Occur Along Island Arcs[J]. Journal of Geophysical Research, 1974, 79(32):4889-4899.

[3] Paul A. Reasenberg, Lucile M. Jones Earthquake Hazard after a Mainshock in California[J]. Journal of American Association for the Advancement of Science, 1989, 243(4895): 1173-1176.

[4] Tibshirani R. Regression shrinkage and selection via the lasso[J]. Journal of the Royal Statistical Society: Series B (Methodological), 1996, 58(1): 267-288.

[5] Friston K J, Frith C D, Liddle P F, et al. Functional connectivity: the principal-component analysis of large (PET) data sets[J]. Journal of Cerebral Blood Flow & Metabolism, 1993, 13(1): 5-14.

[6] Andersen C M, Bro R. Variable selection in regression—a tutorial[J]. Journal of Chemometrics, 2010, 24(11‐12): 728-737.

[7]Zou H. The adaptive lasso and its oracle properties[J]. Journal of the American statistical association, 2006, 101(476): 1418-1429.

[8]Huang J, Ma S, Zhang C H. Adaptive Lasso for sparse high-dimensional regression models[J]. Statistica Sinica, 2008: 1603-1618.

[9] Hoerl A E, Kennard R W. Ridge regression: Biased estimation for nonorthogonal problems[J]. Technometrics, 1970, 12(1): 55-67.

[10] Jolliffe I T. A note on the use of principal components in regression[J]. Journal of the Royal Statistical Society: Series C (Applied Statistics), 1982, 31(3): 300-303.

[11] Geladi P, Kowalski B R. Partial least-squares regression: a tutorial[J]. Analytica chimica acta, 1986, 185: 1-17.

[12] Basler K. Strength of plate girders in shear, Proc. ASCE, 87,(ST7),(October 1961), Reprint No. 186 (61-13)[J]. 1961.

[13] Ali, A. and J. K. Wight. 1990. Reinforced Concrete Structural Walls with Staggered Opening Configurations Under Reversed Cyclic Loading, Report UMCE 90-05. Department of Civil Engineering, University of Michigan, Ann Arbor.

[14] Building Seismic Safety Council. 2000. NEHRP Recommended Provisions for Seismic Regulations for New Buildings and Other Structures, FEMA 368 and 369. FEMA, Washington, D.C.

[15] Bradley, Brendon A. "Correlation of arias intensity with amplitude, duration and cumulative intensity measures." Soil Dynamics and Earthquake Engineering 78 (2015): 89-98.

[16] Campbell, Kenneth W., and Yousef Bozorgnia. "Prediction equations for the standardized version of cumulative absolute velocity as ada pted for use in the shutdown of US nuclear power plants." Nuclear Engineering and Design 241.7 (2011): 2558-2569.

[17] Campbell, Kenneth W., and Yousef Bozorgnia. "A ground motion prediction equation for the horizontal component of cumulative absolute velocity (CAV) based on the PEER-NGA strong motion database." Earthquake Spectra 26.3 (2010): 635-650.

[18] Christopher G. Healey Choosing Effective Colours for Data Visualization[J]. Proceedings of Seventh Annual IEEE Visualization, 1996.

[19] Juha Vesanto SOM-Based Data Visualization Methods[J]. Journal of Intelligent Data Analysis, 1999, 3(2): 111-126.

[20] Andreas BUJA, Dianne COOK, Deborah F. Interactive High-Dimensional Data Visualization[J]. Journal of Computational and Graphical Statistics, 1996, 5: 78-99.

[21] Ying Zhu Introducing Google Chart Tools and Google Maps API in Data Visualization Courses[J]. [IEEE Computer Graphics and Applications](https://ieeexplore.ieee.org/xpl/RecentIssue.jsp?punumber=38), 2012, 32: 6-9.

[22] A. Björklund, N. Gustavsson Visualization of geochemical data on maps: New options [J]. Journal of Geochemical Exploration, 1987, 29: 89-103.

[23] N. Gustavsson, E. Lampio, T. Tarvainen Visualization of geochemical data on maps at the Geological Survey of Finland[J]. Journal of Geochemical Exploration, 1997, 59: 197-207.