
Earthquake-Induced Structural Damage Classification Algorithm

Max Ferguson, maxkferg@stanford.edu, Amory Martin, amorym@stanford.edu
Department of Civil & Environmental Engineering, Stanford University
CS 229 Machine Learning - Final Project

December 12, 2015

Abstract

A classification algorithm is developed for evaluating the damage state of buildings subjected to earthquakes. Nonlinear response history analysis is used to generate the time histories of each building subjected to each earthquake. This report summarizes the analysis procedure used to extract data and describes the different classification algorithms that are developed to predict damage state. Support vector machines (SVM), multinomial logistic regression and k-nearest neighbors (KNN) are considered for the classification algorithm. Features are selected from building parameters, ground motion parameters and combinations of both. Four damage states are estimated: minimal damage, moderate damage, severe damage and collapse. The support vector machine classifier yields the most accurate results for evaluating the damage state with an overall accuracy of 71.2%. This is in part due to the highly nonlinear nature of this problem. The accuracy of classifying a damage state with a misclassification tolerance of ± 1 is 95.2%.

I Introduction

Chile (2010), Tohoku (2011) and Christchurch (2011) were particularly devastating earthquakes. Because these natural disasters are very difficult to predict, they often come with a high death toll. Today, structural engineers design stronger and more earthquake-resilient buildings thanks to the development of nonlinear dynamic analysis. However, post-earthquake evaluation of structures is still a difficult task. Quantifying the damage of a given structure is both technically challenging and time-consuming. Only certified experts can use their judgement to evaluate earthquake damage. A damage state classifier for several damage states is developed. Support vector machine (SVM), multinomial logistic regression and k-nearest neighbors (KNN) were considered for the classification algorithm. The classifier predicts the post-earthquake damage state, given the building features and input ground motion. The ultimate purpose of this classifier is to accelerate post-earthquake damage evaluation of critical buildings to allow a faster recovery and decrease financial losses expected from downtime and repair.

Within the field of earthquake engineering there have been numerous attempts to develop a closed-form equation to classify the displacement and correspond-

ing damage state of a multistory building subject to a particular ground motion (Miranda, 1999; Stephens & Yao 1983). Past research in this field tends to focus on either quantifying the intensity of a particular ground motion or relating structural response to ground motion parameters (Gupta & Krawinkler, 2000). However, it is proposed that machine learning techniques can be applied to develop a classifier that predicts earthquake damage based on a wide range of building and ground motion features.

II Data Extraction

Nonlinear response history analysis is performed to generate data for the classification algorithm. Using a wide range of parameters, such as mass and stiffness for each story, 300 theoretical buildings are modeled. Using the Pacific Earthquake Engineering Research (PEER) data, 200 ground motions are selected and scaled appropriately. Nonlinear dynamic analysis is performed on every building to create roughly 60,000 displacement time histories. This process is both time consuming and challenging, given the computationally intensive nature of nonlinear dynamic analysis.

Building Parameters

A set of 300 theoretical structures were used to represent common building types. Each building has a unique set of structural parameters. The following structural parameters were considered: (1) mass M , (2) stiffness K , (3) fundamental period of vibration T_n , (4) number of stories n , (5) height h , (6) damping ratio ξ , and (7) yield strength F_y .

Ground Motion Records

A set of 200 ground motion time histories were obtained from the Pacific Earthquake Engineering Research (PEER) data. Each ground motion time history contains the recorded lateral ground acceleration in two perpendicular directions. Therefore, a total of 200 acceleration records were available for generating data. The Fourier spectrum of each record is calculated by transforming the acceleration record into the frequency domain as shown:

$$A(f) = \int_{-\infty}^{+\infty} A(t)e^{-2\pi ift} dt$$

Using both the Fourier spectrum and time record, the following parameters are obtained from each ground motion: (1) peak ground acceleration PGA , (2) the maximum Fourier acceleration amplitude FA_{max} , (3) the Fourier acceleration amplitude at a period of 1 second FA_{1s} , (4) the dominant period of the Fourier spectrum and (5) the total earthquake energy defined as $E_A = \int A(f)df$. These parameters are chosen to represent both the severity and frequency content of the earthquake.

Equivalent SDOF Model

To extract a large number of data points the structural models are reduced from a multi-degree of freedom (MDOF) system to a single-degree of freedom (SDOF) system. Figure 1 shows the equivalent model. The equivalent stiffness, mass and height SDOF parameters are obtained using the following equations:

$$H_e = \frac{\sum_{i=1}^n m_i h_i^2}{\sum_{i=1}^n m_i h_i^2} \quad M_e = \frac{\sum_{i=1}^n m_i h_i^2}{H_e}$$

$$T_e = 2\pi \sqrt{\frac{M_e}{K_e}} \quad K_e = K_{eff} H_e^2 \quad K_{eff} = \lambda K_s$$

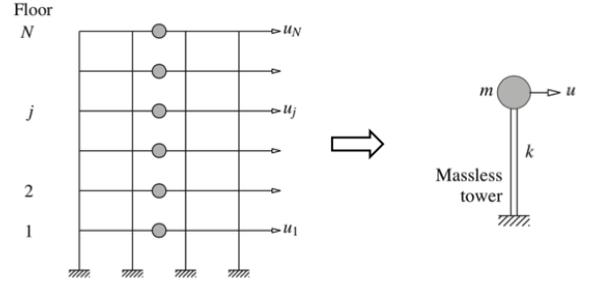


Figure 1: Equivalent single degree of freedom (SDOF) structural model

Response History Analysis

A response history analysis (RHA) is performed on each building-ground motion combination to obtain a time history response. The RHA computes the displacement, velocity and acceleration response of a structure to a ground motion input. The displacement of the structure is the solution of a system of 2^{nd} order differential equations

$$M\ddot{u}(t) + C\dot{u}(t) + Ku(t) = -M\ddot{u}_g(t)$$

$$\Phi^T M \Phi \dot{q}(t) + \Phi^T C \Phi \dot{q}(t) + \Phi^T K \Phi q(t) = -\Phi^T M \ddot{u}_g(t)$$

Newmark's average acceleration method is chosen as the direct integration scheme ($\gamma = 1/4$, $\beta = 1/2$) with a Newton-Raphson algorithm for the nonlinear convergence. The maximum displacement, inter-story drift, roof acceleration and base shear force are obtained from the time histories. In earthquake engineering, out of these parameters, the peak inter-story drift has been shown to directly correlate with the damage of the building.

III Method

Features from the structure, the earthquake and a combination of both are used for the classification algorithm. To obtain some of the ground motion features, a Fourier analysis is conducted. Additionally, using a moving average smoothing function and K-means clustering, a ground motion type is determined and used as one of the features. Each classification algorithm is first trained then tested using cross validation of the obtained data. The damage state is therefore predicted for each (structure, earthquake) combination. Figure 2 shows the flowchart of the general approach.

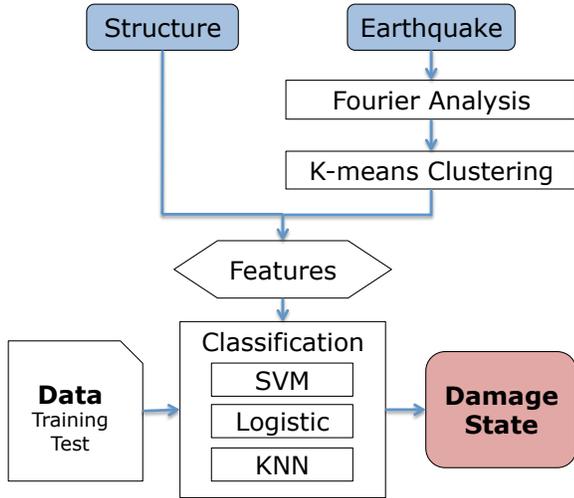


Figure 2: Flowchart

Damage States

A damage state is the level of structural damage of a given building in a post-earthquake scenario. In earthquake engineering, the damage state is directly correlated to downtime, cost of repair and financial losses. Fragility functions are used to identify damage states for given structures. For the purposes of this report, four damage states were identified:

1. DM_1 : superficial damage
2. DM_2 : moderate damage
3. DM_3 : severe damage
4. DM_4 : collapse state

Superficial damage corresponds to non-structural damage such as cladding, partitions and glass damage. Moderate damage consists of limited structural damage to beams and columns. Severe damage corresponds to damage to main lateral systems and residual drift. Usually these buildings have a residual drift and are highly susceptible to collapse during subsequent aftershocks. Finally, collapse state is the last damage state where the building is considered highly unstable. The building is evacuated immediately and remains closed until significant repair or demolition.

Earthquake Frequency Categorization Using K-means Clustering

Ground motion records tend to fall into one of several categories, based on the magnitude, local soil type, and fault distance. Figure 3 shows the earthquake Fourier spectra.

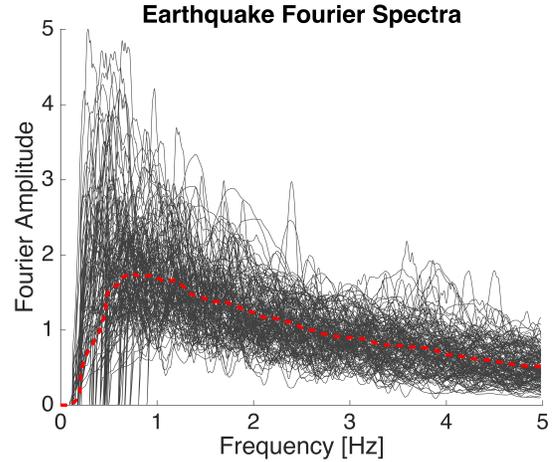


Figure 3: Fourier spectra of sample ground motions

Each ground motion category has a relatively distinct frequency content. The k-means clustering algorithm, along with a smoothing moving average, is used to group the earthquake records into six different types based on the Fourier spectrum. Figure 4 shows the resulting Fourier spectra for each earthquake type. The ground motion type is used as a feature in the damage classifier, as the dynamic response of a structure is correlated with the frequency content.

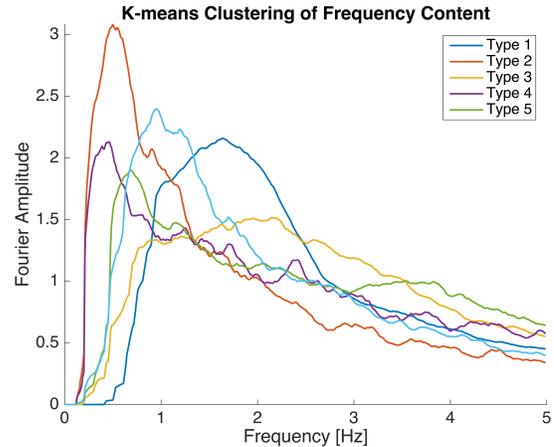


Figure 4: Earthquake type using k-means clustering

Features

A range of features describing both the building and the earthquake are used to train the classifier. Furthermore, nonlinear combinations of these features are derived and used as additional features. Most of these features are commonly used in earthquake engineering. In theory, the predicted damage state is not proportional to individual features. In fact, damage state depends on a complex nonlinear combination of these features. For instance, surprisingly, there is no obvious linear relationship between PGA and damage state. The following is a non-exhaustive feature list:

Building Features

1. Mass M
2. Stiffness K
3. Fundamental period of vibration T_n
4. Number of stories n
5. Height h
6. Damping ratio ξ
7. Yield strength F_y

Ground Motion Features

8. Peak ground acceleration PGA
9. Maximum Fourier amplitude FA_{max}
10. Fourier amplitude at a period of 1s FA_{1s}
11. Dominant period of the Fourier spectrum T_A
12. Total earthquake energy E_A
13. Earthquake type

Derived Features

14. Period shift $(T_n - T_A)^2$
15. Energy absorption $(T_n - T_A)^2 PGA$
16. Dissipation factor K/PGA

Support Vector Machine Classifier

A support vector classifier is developed to predict the damage state of a structure, given a range of building and ground motion features. The multiclass classifier is composed of three independent binary SVM classifiers. The first predicts whether the structure will be lightly damaged or severely damaged. The second predicts the damage state of a lightly damaged structure, while the third predicts the damage state of a heavily damaged structure. Each classifier is optimized and trained individually to ensure that the overall accuracy is high. Multiple kernels are tested, with the radial basis function kernel (Gaussian kernel) providing the best results. A parameter grid search with 5-fold cross validation is used to select the optimum C and γ values for each binary classifier.

Multinomial Logistic Regression

A multinomial logistic regression classifier is also developed using polytomous responses for estimating the damage state. A cross validation of 70/30% is used training and testing the data. An ordinal model with interaction between the multinomial categories and coefficients is implemented. Using the statistical significance of the model coefficients, the best performing features are selected and chosen for estimating the damage state.

Nearest Neighbors Classifier

A nearest neighbors classifier is developed to predict the damage state of a particular structure after being subjected to a ground motion. To achieve adequate performance the number of training data points is increased to 540,000. A classifier is developed which selects the closest five scenarios from the dataset using the Euclidean distance as the distance metric. The classifier returns the damage state that occurs most frequently across the five neighbors. If two or more damage states are equally probably, then the largest is selected.

IV Results

Three classifiers were developed to predict the damage state of a structure subject to a particular ground motion. Figure 5 demonstrates how the accuracy of each classifier was related to the number of training features.

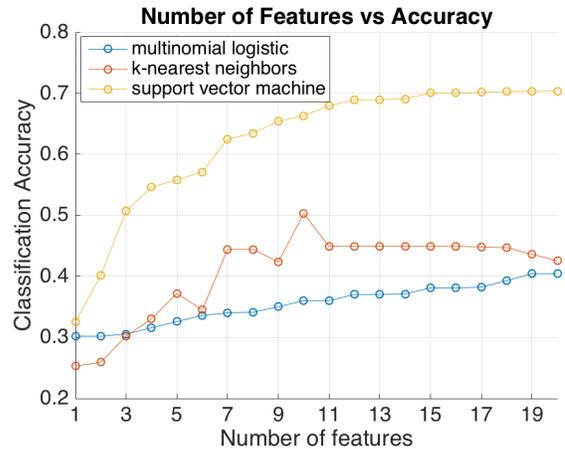


Figure 5: Classification accuracy of each classifier

The accuracy of the SVM algorithm monotonically increased as more features were added to the model. The accuracy of the multinomial logistic classifier was relatively unchanged by the number of features, demonstrating that the shape of the damage-state surface is difficult to capture with the linear combination of features. Finally, K-nearest neighbors (KNN) increased parabolically as the number of features grew.

Overall, the multinomial logistic regression was the worst classifier compared to the SVM and K-nearest neighbors classifier. Due to highly nonlinear trend, this is expected. Using a nonlinear combination of features improved the accuracy of the result, however, determining which combinations to use is time-consuming and challenging. Moreover, when too many features are used, the multinomial logistic regression has difficulties converging; as n simultaneous equations

are solved for n features, the conditional number can be poor.

		Predicted Damage State				Total	Recall
		1	2	3	4		
Actual State	1	3,353	2,756	942	1,631	8,682	0.386
	2	1,457	7,197	1,754	1,719	12,127	0.593
	3	410	1,455	5,903	2,701	10,469	0.139
	4	414	1,595	3,690	6,583	12,282	0.536
Total		5,634	13,003	12,289	12,634	43,560	
Precision		0.595	0.553	0.480	0.521		

Figure 6: Confusion Matrix for KNN Classifier

The k-nearest neighbors classifier showed a relatively high accuracy when trained and tested with 10 features. The addition of more features did not improve accuracy. The results shown in Figure 6 were generated using 540,000 training points. Classification accuracy can be improved by adding more training points, but this would make the classification of new points very computationally demanding.

		Predicted Damage State				Total	Recall
		1	2	3	4		
Actual State	1	6,980	2,335	270	97	9,682	0.721
	2	795	8,165	1,865	302	11,127	0.734
	3	234	2,502	6,894	839	10,469	0.659
	4	110	1,063	2,132	8,977	12,282	0.731
Total		8,119	14,065	11,161	10,215	43,560	
Precision		0.860	0.581	0.618	0.879		

Figure 7: Confusion Matrix for SVM Classifier

The SVM classifier provided an impressive 71.3% classification accuracy on the testing data. The precision and recall for each damage state were also relatively high, when compared to the other methods. The accuracy of classifying a damage state with a misclassification tolerance of ± 1 was 95.2%

V Conclusion

A large set of simulated data was collected using dynamic nonlinear response history analysis. Ground motion features were determined using moving average functions, Fourier analysis and K-means clustering. Structural features and earthquake features were combined to form additional features. The key features were period shift, ground motion energy, peak ground acceleration (PGA), stiffness as well as nonlinear combinations of these such as energy absorption. Three classification algorithms were used and

tested using cross validation: (1) multinomial logistic regression, (2) K-nearest neighbors (KNN) and (3) support vector machine (SVM). Due to the nonlinearity of ground motion response, the SVM outperformed the other two classifier with a 71.3% overall correct classification. The performance of the SVM classification algorithm is impressive, given the complexity of dynamic structural analysis. In earthquake engineering, the damage state is often contested and defined using different fragility functions. Small misclassification is not of great importance. For instance classifying a building as DM_3 severe damage instead of DM_2 moderate damage is not a crucial mistake. The accuracy of classifying a damage state with a misclassification tolerance of ± 1 was 95.2%.

VI Further Work

Further work is possible for predicting damage state of buildings and in other engineering applications of machine learning. For instance, damage state classification algorithms can be developed for nonstructural elements. Machine learning techniques have a wide range of applications in structural engineering. Using the derived damage state classification algorithm, community and city-wide expected financial loss estimation based on earthquake probabilities in the region can be determined. Finally, machine learning techniques can be applied to control problems. For instance, active base-isolation systems can be designed using machine learning for optimal structural control.

References

- [1] Anil K Chopra. *Dynamics of Structures*. Prentice Hall, Fourth Edition 2011.
- [2] Chang, C. and Lin, C. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*.
- [3] Gupta, A. & Krawinkler, H. (2000). "Estimation of seismic drift demands for frame structures". *Earthquake Engineering and Structural Dynamics*.
- [4] Jayaram, N., Lin, T., and Baker, J. W. (2011). "A computationally efficient ground-motion selection algorithm for matching a target response spectrum mean and variance." *Earthquake Spectra*.
- [5] Miranda, E. (1999). "Approximate Seismic Lateral Deformation Demands in Multistory Buildings." *J. Struct. Eng.*
- [6] Stephens, J. and Yao, J. (1987). "Damage Assessment Using Response Measurements." *J. Struct. Eng.*