Neural networks explanation models to support earthquake mitigation decision-making

Rodrigo Silva-Lopez
Blume Earthquake Engineering Center
Stanford University, USA
rsilval@stanford.edu

1 Introduction

Rapid urbanization and climate change have made cities more vulnerable to the occurrence of natural disasters. Earthquakes are among the natural disasters than severely impair communities, with the occurrence of the events in Haiti 2021 and Japan 2011 being vivid examples of the ability of these events to disrupt our lives. Motivated by minimizing the risk that earthquakes impose, researchers have developed complex models to accurately determine the impacts of these events, allowing to quantify the effect of mitigation actions. These earthquake models are computationally expensive, challenging the ability of researchers to translate their results into public policy. It is in that regard that neural networks have risen to generate surrogate models that rapidly and accurately can estimate the impacts of earthquakes. However, there is significant skepticism in disaster science to use neural networks due to their lack of interpretability. Considering the previous challenges, this study aims to develop a neural network to rapidly and accurately predict recovery time of buildings that experience damage after an earthquake. Moreover, besides just calibrating a neural network, this project implements explanation models to understand the internal mechanisms of neural network prediction and propose mitigation actions based on these results.

2 Task developed by the neural network

The task developed by the neural network is to predict the median recovery time for a given structure subject to a specific set of ground motion intensities. Recovery time of a building is defined by the strength of its non structural components such as partition walls, ceilings or elevators. The building recovery model used in this study considers 41 of these components. During the computation of recover for a given seismic scenario, each of these components is damaged according to their strength, defined by a lognormal CDF function. In terms of decision variable, structural engineers can define the value of the median that defines the lognormal CDF. Considering this, to replicate the decision process, the neural network uses as an input a multiplying factor α for each non-structural component k. Following what is done in practice, α has a range between 1 and 3. This α factor multiplies the original median assigned to the component k according to existing building codes. Considering the above, the calibrated neural network in this study uses an an input a vector of 41 values of α, one for each component, and it provides as output the value of median recovery time obtained from 1000 realizations of running a recovery model.

3 Dataset generation

As a first step of the calibration of the neural network, data was generated using the recovery model developed by (Cook, 2021) for a three story building located in the city of Oakland. For each combination of α of the components, 10000 runs of the recovery model were obtain, each of them had different realizations of damage, which were determined by the probabilistic nature of using a lognormal CDF as a fragility function to sample damage. This work considers 10000 realizations for each input vector since it proved to be stable. In terms of data generation, for each set of 10000 realization, a single output corresponding to the median value was obtained. To train the neural network, 20000 values of median recovery times, each having different combinations of α factors

were obtained. Each combination was sampled uniformly randomly from a range between 1 and 3 as mentioned above. Given the computational costs involved in this data generation, to facilitate the production of these values, the High Performance Computing facility at Stanford University, Sherlock, was used.

4 Approach

4.1 Overview

As presented in the introduction, this work has two main steps: (1) Neural network training and (2) Neural network explanation. The first step has as an objective to develop an accurate neural network to predict recovery time. The second step aims to understand the neural network to propose mitigation actions. More details about each of these steps are presented next.

4.2 Neural network training

The first step of this work is to calibrate a neural network to accurately predict median recovery time. To that purpose, a hypercalibration is performed. Accuracy of the neural network is measured in terms of coefficient of variation R^2 and bias is measures by taking values of R^2 across different windows of values of recovery time. Using these metrics, the parameters with highest accuracy on the validation data are selected from the hyperparameter calibration. These results are shown in the section of Results.

4.3 Neural network explanation

The second step involved in this work is to implement neural network explanation models to motivate mitigation actions. In particular, this study detects what components seem important for the neural network and explore how improving the performance of those component can decrease recovery time of buildings. Importance in the neural network for effects of this study is measured in two ways, (1) as the change of accuracy of the neural network generated by dropping the component from the training, and (2) also importance is computed using LIME algorithm. The effects of retrofitting those components is measured by the recovery model and by comparison with existing understanding of structural components. To initiate the experiments aimed at developing a deeper understanding of the neural network, a counterfactual validation is performed.

4.3.1 Counterfactual validation

A counterfactual validation aims to evaluate the capacity of the neural network to predict important data points, as defined by the users (in this case structural engineers), without having observed those cases before. By having the capacity of doing so, the neural network shows that it is not only interpolating between known information, but it is also learning about the underlying model, developing the ability to extrapolate its knowledge. In this study, the contractual information considers a comparison with a sensitivity analysis in which each component is retrofitted individually, and also the use of edge cases in which some components are retrofitted jointly based on their similarity as defined in building codes. In addition, based on conversations with experts, this study uses as important data points to calibrate, those in which non-structural components are retrofitted at the same time according to their structural group. For instance, all components that correspond to partitions are retrofitted simultaneously in these data points.

By doing this conterfactual validation, this study can prove to skeptics in civil engineering that this tool has a powerful capability of understanding the underlying model and not being just a non-linear regression, further validating the use of this neural network.

4.3.2 Simple explanation model

The first explanation model, named here as simple explanation model, computes the importance of each component based on the change of accuracy of the neural network when the component is excluded during the training process. The bigger the decrease of accuracy on test data, then the more important the component is.

4.3.3 LIME explanation model

The second importance variable method implemented is the model developed by (Ribeiro et al., 2016), in which each input variable has assigned a global importance weight based on their local weight associated to local linear regressions. The bigger the weight is associated to a component, the more important it is.

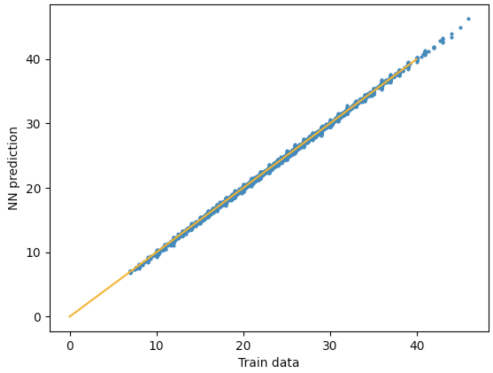

Figure 1: Scatter plot showing results for training data

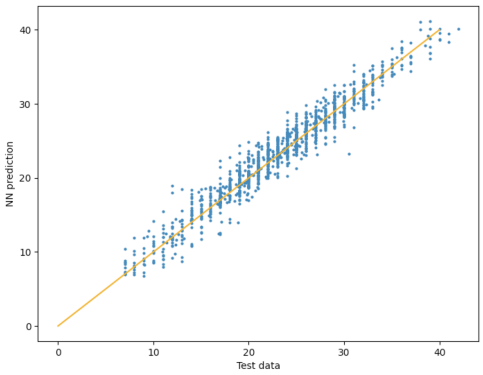

Figure 2: Scatter plot showing results for the test-set

4.3.4 Proposal of mitigation actions

To validate that the results of the variable importance method are appropriate, a comparison with expert criteria is performed.

5 Results

5.1 Neural network training

In this section we present the results of the calibration of the neural network on the training and test set. In the training set, the value of R^2 was 0.99, and it reached 0.95 on the test set. Simpler regression techniques such as linear regression, Lasso regression and random forests yielded performance of 0.65,0.71,0.80 respectively, showing that the use of neural networks is appropiate in this case.

The plot for the test set is shown in Figure 2

Some of the hyperparameters obtained through the hyperparameter calibration are shown in Table 1.

Table 1: Neural network hyperparameters

Hyperparameter	Value
Number of layers	20
Neurons per layer	150
Learning rate	0.0003

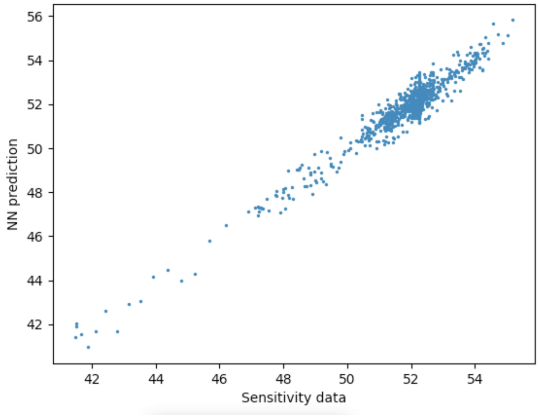

Figure 3: Scatter plot showing the results of a sensitivity analysis computed by using the neural network versus using the recovery model

5.2 Neural network explanation

5.2.1 Counterfactual validation

- Sensitivity Analysis

  As part of the counterfactual validation, the results predicted by the neural network were compared with performing a sensitivity analysis in which each component improved its value of α individually and in steps of 0.1. Note that since the neural network is trained using a random combination of α_k the specific values used in the sensitivity analysis have not been directly observed by the model.

- Performance on important data points

  In addition to the sensitivity analysis, we evaluated the performance of the neural network at predicting Important Data points (ID) as defined by experts in building recovery. These points comprise (1) Retrofitting components simultaneously if they belong to the same structural group (2) Incrementally retrofitting all components at once, and (3) Retrofitting one component at a time, similarly to the sensitivity analysis shown before. The results of this analysis are presented in Table 2, where "Random Sampling" corresponds to the protocol in which

each α for each components was generated randomly, and is the approach shown in previous sections. The performance of this approach is measured in terms of R^2 for three datasets of interest: (1) All data, which includes random data and Important Data points, (2) Important Data points (ID), and (3) Other data points, which correspond to random data points. We observe that while using random sampling ensures an overall successful performance, it does not show a high performance for Important Data points. Motivated by this, we explored two additional training sampling protocols, one in which some ID points were added to the training set, and other in which the training process was only performed in these ID points. As expected, adding the ID points significantly improved the performance at predicting ID points while keeping and overall performance stable. On the other side, just training on ID points performed well on ID points, but showed a poor performance on the rest of the dataset.

5.2.2 Variable importance algorithms

Two variable importance algorithms are used in this study. The first one ranks each variable by the decrease in performance associated to taking that variable out of the training process, and the second one is the implementation of submodular LIME. The results of these implementations are shown in Table 3, which show elements that generate a change in R^2 of more than 0.01, considered as meaningful given the uncertainty of the model. Both variable importance models match respect to the ranking they provide. They also match the results of using the recovery model with a one-at-a-time analysis, which gives credibility to the obtained results.

5.2.3 Performance of the neural network as more variables are included

Despite 41 variables are included to generate recovery times, some of the components do not seem to play an important role in the estimation of recovery time. Motivated by this, we used the structure of the neural network as a proxy to define what components should be considered and which ones should be neglected. To achieve this, we trained the neural network by incrementally adding variables in the dataset. The order in which the variables were added was consistent with the ranking obtained from the variable importance algorithms. The results of this experiment are shown in Table 4, where we observe that after adding the first 10 variables the performance of the neural network at estimating recovery times is equivalent to including all 41 variables in the analysis, therefore the rest of the 31 variables may not be necessary to estimate recovery times, and from the perspective of policy making, those components should not be targeted.

5.2.4 Proposal of mitigation actions

The results of the previous sections on Variable Importance and Evolving Performance can provide insights on potential mitigation actions to decrease recovery time of buildings. From the Evolving Performance we can confirm which components do not have an impact on evaluating recovery times. Moreover, given that the results of both Variable Importance algorithms is similar, we can conclude that the proposed ranking provides an order in which the components could be retrofitted to minimize recovery.

In addition to the previous proposal of mitigation actions, we established conversations with experts in the field, who agreed with some of the most important components identified by the neural network such as Partitions, Curtain walls, Prefabricated stairs and suspended ceilings. Therefore, the proposal of these elements to be retrofitted is consistent with expert criteria.

6 Future work

This work showed the potential of using surrogate models to improve the seismic recovery assessment of buildings. As future work, other building configurations and locations will be explored. In particular, we are interested in the performance of these systems for complex building of several stories where recovery computations are particularly slow, therefore it will be critical to analyze how much data can be used as a minimum threshold to ensure performance. Once using these surrogate models for several buildings and sites has been validated, we will explore regional models to analyze public policies aimed at improving the resilience of our communities.

Table 2: Performance of neural network for different training protocols and datasets

Sampling Protocol	All data	Important Data points (ID)	Other datapoints
Random	0.959	0.713	0.942
Random + ID	0.955	0.945	0.944
ID	0.021	0.951	0.001

Table 3: Implementation of variable importance algorithms

Ranking	Component	Decrease in R^2	LIME value
1	Steel stair	0.6361	345.66
2	AC Drops	0.4461	234.98
3	Curtain wall	0.410	226.74
4	Pendant Lighting	0.246	115.78
5	Suspended Ceiling	0.119	59.34
6	Partitions	0.064	27.65
7	Suspended Ceiling II	0.035	25.91
8	Cladding Panels	0.019	12.54
9	HVAC Fan	0.018	10.23
10	Water Piping	0.017	9.87

Table 4: Evolving performance of neural network by adding variables

Components	Component	R^2
1	Steel stair	0.021
2	AC Drops	0.319
3	Curtain wall	0.489
4	Pendant Lighting	0.581
5	Suspended Ceiling	0.841
6	Partitions	0.857
7	Suspended Ceiling II	0.893
8	Cladding Panels	0.901
9	HVAC Fan	0.933
10	Water Piping	0.958

7 Conclusions

This study introduces a neural network-based surrogate model to estimate building recovery after earthquakes. The model proved to be around 10,000 faster while still obtaining accurate results. Besides overall performance, the calibrated neural network proved to be able to predict some edge cases such as a sensitivity analysis and Important Datapoints (ID). However, to improve the performance on the later data it is important to include some of them into the training set.

In addition to the calibrated neural network, this study implemented variable importance algorithms which led to the identification of the components that drive recovery time on buildings. Both variable importance methods showed the same results, and the exclusive inclusion of the variables that were tagged as important led to an accuracy of the neural network similar to using all the variables, therefore validating these components as the ones that define the response of the buildings.

Finally, the proposal of important components can lead to the proposal of mitigation actions by identifying non-structural components that can have a greater impact at improving the performance of our built environment. The proposed order of components is consistent with expert criteria in the field, further validating the results and the potential of using neural network-based surrogate models.

8 Contributions

This work was mostly developed by Rodrigo Silva-Lopez. Omar Issa, a Master's student working in the research group of Professor Jack Baker, will continue using these models as part of his research, therefore he was the one that raised the need to develop this model and helped to develop data that adjusted to his needs.

References

Dustin Trevor Cook. 2021. *Advancing Performance-Based Earthquake Engineering for Modern Resilience Objectives*. Ph.D. thesis, University of Colorado at Boulder.

Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. " why should i trust you?" explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1135–1144.