Visualizing Earthquake Simulation Data

Zhenge Zhao,* Youhao Wei,† Joshua A. Levine,‡ Matthew Berger,§ Danilo Motta,¶ Carlos Scheidegger‖

ABSTRACT

Computer simulations of the effect of earthquakes on built structures promise to let engineers understand different tradeoffs and designs at an attractively low cost. In this scenario, the bottleneck for the expert is one of *data understanding*: how do different building designs respond to different earthquakes? What do the building failure modes have in common, and how does this compare to theoretical predictions? The way in which a building responds to an earthquake is complex and controlled by different factors. In this poster, we present ongoing work in building a system for interactive visualization of earthquake simulation data, in collaboration with civil engineers who have run thousands of simulations, varying the height of the simulated buildings, ground acceleration, and building structure design. We describe the challenges in the visualization of such multivariate time series data, and some of our proposed solutions.

Index Terms: Visualization, Multivariate time series, Earthquake Simulation, NLP, Kernel

1 INTRODUCTION

Through computational simulation, civil engineers can understand how buildings respond to external forces much more efficiently than previously possible. The simplicity with which different scenarios can be simulated and analyzed is attractive, but it is ultimately the source of a new problem: the understanding of the phenomenon is now dependent on the ability to quickly make sense of large amounts of data.

In this poster, we will report on an ongoing collaboration with civil engineers who run a large number of simulations of building responses, and how data analysis and information visualization can highlight interesting patterns in the data. This problem is especially challenging because of the variety of scales involved. There are multiple earthquakes to be compared to each other; each building's response to an earthquake naturally varies across simulation time, and has different values along different points of the building. There are also periodic phenomena in each simulation, occurring at potentially different periods. Finally, there are multiple physical variables of interest, including shear, moment, and diaphragm forces.

Some of the visualization techniques we using include matrix diagrams and multivariate time series visualizations, as can be seen in Figure 1. We combine those techniques with classical techniques from signal processing for segmentation, and recent techniques in machine learning for determining similarities between segments of each earthquake simulation.

2 RELATED WORK

Although the problem of understanding how a building behaves during an earthquake has long been studied. In order to investigate the performance of a building on a shake table, in 2007, engineers built a seven-story building for testing and re-creation of a seismic response of it based on recorded data of a full scale shake table test [3]. Here, we are concerned about the relationship between the different physical measurements, as well as with the understanding of how these patterns are similar or different across different earthquakes.

At the same time, visual exploration of multivariate data sets is an integral part of scientific visualization. As in most real world phenomena, there exist multiple factors associated with the complex interactions of different variables. To gain an in-depth understanding of a scientific process, the relationship among the variables needs to be thoroughly investigated. Biswas et al. [1] propose a framework to classify isocontours of variables based on the relationship between them and users can explore the multivariate data sets using their interface. Comparing time series data is another great challenge. Kernel methods, such as Support Vector Machines and Gaussian Process have become classical data-analysis tools. Vert et al. propose an alignment kernel for time series which is widely used when comparing time series [4].

3 ONGOING WORK

Our ultimate goal is to build infrastructure to enable civil engineers to visually compare these multi-variate temporal data. In collaboration with the domain experts, we arrived at the following visualization tasks:

- T1 An overview which summarizes all the earthquake simulations and can help them spotting outliers;
- T2 For a specific earthquake simulation, a context view to explore the multivariate dataset and help engineers find interesting patterns across time;
- T3 For the two views above, build efficient navigation and interaction between them;
- T4 When an interesting event is identified, it should be possible to find similar events on other simulations, so the context in which these events happened can be compared.

Towards this end, we built a prototype interface using 50 of the available earthquake simulations. The simulations track 6 physical attributes of interest, for each floor: acceleration, shear, diaphragm force, moment, drift ratio, and interstory drift ratio. For example, shear measures the stress parallel to each floor, while interstory drift ratio measures the positional difference between two adjacent floors at a given point in time. This gives a vector-valued time series for each attribute, and each simulation has $25,000$ time steps in average. Each attribute is normalized by dividing the raw value by a predetermined design limit. This has the benefit that any value of the time series above 1 or below negative 1 indicate that the building is operating out of its safe design specifications, and mitigates the issue of comparing variables of different units.

Implementation details The current prototype is implemented in Javascript using D3 [2], and employs both SVG and HTML5 Canvas for performance. The earthquake simulations are run on a cluster, and the output is preprocessed using Python and SciPy before they are consumed in our D3 application.

*zhengezhao@email.arizona.edu
†youhaowei@email.arizona.edu
‡josh@email.arizona.edu
§matthewberger@email.arizona.edu
¶ddanilomotta@gmail.com
‖cscheid@email.arizona.edu

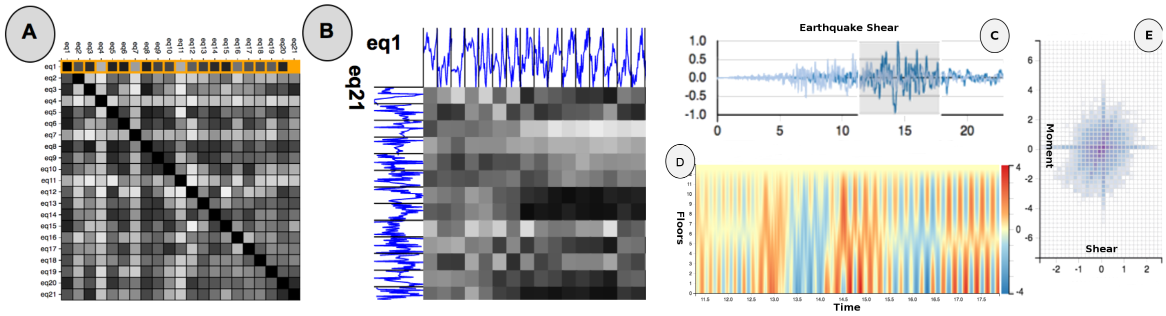

Figure 1: Some of the multiple views supporting comparative visualization of earthquake simulations. The response of a building for the entirety of the simulation can be compared to one another, and some earthquakes have more similar responses to others. This gives a comparative overview of all simulations in (A), where a matrix diagram view shows the overall similarities of floor shear across 50 earthquake simulations. In addition, building responses tend to be vibrational and periodic in nature, so segments of an earthquake simulation are compared to one another in (B), where a matrix diagram view demonstrates the similarities between different segments of two earthquake simulations (eq1 and eq21 in view (A)). Analysts can select a portion of the ground acceleration (C) and drill down into a specific earthquake simulation (D), to visualize the response of a single physical variable plotted over time (x coordinate) and building floor (y coordinate). Finally, a 2D histogram can be used to compare two different attributes over the same period of time: how does shear (x coordinate) compare to moment (y coordinate)?

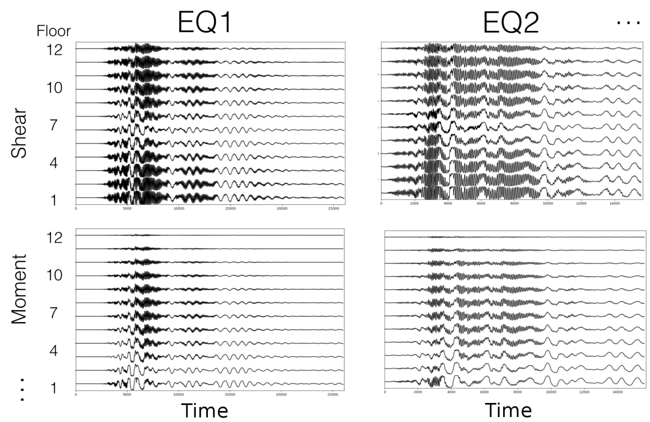

Figure 2: An overview of the data collected by the civil engineers during the simulation. The main challenge in this project is to design visualizations which highlight interesting patterns across the different dimensions, attributes, and scales.

NLP Model and Probability Product Kernel for cross-earthquake comparisons A different view (left part of Fig. 1) organizes the entire dataset more abstractly. For example, suppose we want to compare the shear value across different earthquakes. Based on the fact and observation that simulation data has periodic features, we use a bag-of-words analogy inspired from Natural Language Processing (NLP). We split each earthquake simulation into a set of segments with a fixed period (P), producing segments that are represented as $13 \times P$ matrices. After applying the segmentation, each earthquake simulation becomes a set of matrices (the "words" in the bag of words, which we call "motifs"). For any two motifs coming from two earthquakes (of possibly different periods), we compare them directly as continuous time series, by upsampling the shorter motif to the length of the longer. After representing simulations as motifs, we calculate the similarity between any two earthquake simulations by mapping the set of motifs to a Gaussian distribution in Hilbert space, and use Bhattacharyya's similarity to compare the earthquakes [5].

Matrix diagram views We use matrix diagrams to visualize both the behavior across earthquakes, and within an earthquake, and use D3's existing matrix diagram infrastructure. Each cell in the across-earthquake matrix represents the similarity between two entire simulations using Bhattacharyya's measure, as described above. In this view, users can select cells in order to show a more detailed comparison of one single earthquake against another (or even against itself), comparing all motifs in one earthquake to all motifs in the other. These two views allow users to explore interesting finding. For example, in Fig. 1A, the eleventh earthquake is clearly quite dissimilar to all other earthquakes (as evidenced by the mostly-white row of values).

Ongoing work Even with the matrix diagram visualizations, the screen becomes cluttered as the number of simulation gets larger. In the future, when we get thousands of simulations, the matrix diagram will not scale. It is also worth considering if it's possible to cluster earthquakes hierarchically to maintain visual scalability. The way we choose periods for segmentation is also clearly inappropriate; we are currently investigating how to enable our machine learning methods with multiple overlapping motifs. Finally, even though we have worked with domain experts in developing these tools, a thorough validation of the designs remains to be done.

References

[1] A. Biswas, S. Dutta, H.-W. Shen, and J. Woodring. An information-aware framework for exploring multivariate data sets. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2683 – 2692, Dec. 2013. doi: 10.1109/TVCG.2013.133

[2] M. Bostock, V. Ogievetsky, and J. Heer. D3 data-driven documents. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2301–2309, Dec. 2011. doi: 10.1109/TVCG.2011.185

[3] A. Chourasia. Digital re-creation of a seven-story building shake during an earthquake. *Crossroads*, 13(3):5–5, Mar. 2007. doi: 10.1145/1247238.1247243

[4] M. Cuturi, J. Vert, O. Birkenes, and T. Matsui. A kernel for time series based on global alignments. *CoRR*, abs/cs/0610033, 2006.

[5] R. Kondor and T. Jebara. A kernel between sets of vectors. In T. Fawcett and N. Mishra, eds., *ICML*, pp. 361–368. AAAI Press, 2003.