Project Work 2023/24 I. semester
Timber building modelling

December 17, 2023

David Apagyi
apagyidavid@student.elte.hu

Advisors:
Noemi Friedman (SZTAKI) and Andras Urbanics (SZTAKI)

Introduction

This semester I have been working on an engineering research project at SZTAKI as part of the project work course. The aim of the project is to better understand, analyse and model the behaviour of multi-storey timber buildings. These buildings can exhibit behaviours (e.g. high levels of vibration) that are not common in conventional buildings. The main project related paper [1] addressing this issue presents a surrogate model in place of an existing finite element model for a specific building. My work has addressed the difficulties encountered in constructing a specific part of this model.

Problem statement

The data we work with consists of six parameters and their associated eigenfrequencies and mode shapes, derived from measurements with a total of 10,000 records. The parameters describe six physical properties of the building. For each dataset there are six eigenfrequencies and six associated mode shapes. The eigenfrequencies are positive real numbers, while the mode shapes are 26-dimensional vectors describing the x and y directional vibrations measured at 13 points in the building.

The problem is that due to the overlapping distribution of eigenfrequencies, it is not always correctly identified which measured mode shape actually belongs where. For example, as shown in Figure 1, the measured data for the first three mode shapes are mixed together, instead of the distinct three mode shapes shown on the right. This is because the original ordering relied heavily on the eigenfrequencies, but their distributions overlap greatly, as shown in Figure 2. Our goal is to clean these up as much as possible.

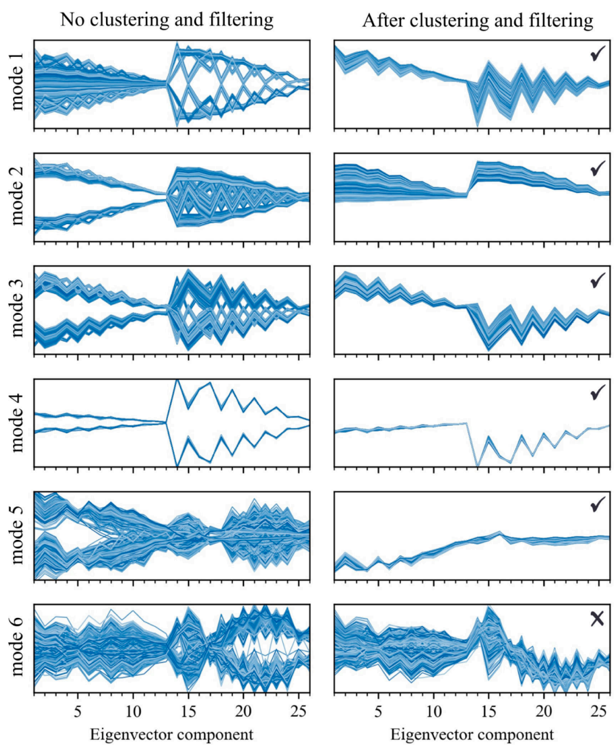

Figure 1: Illustration of the problem of mode shapes [1].

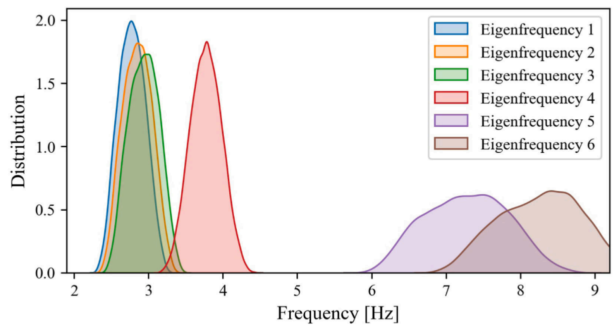

Figure 2: Overlapping distribution of eigenfrequencies [1].

Achievements

The problem can be split into two parts and the data associated with the first three mode shapes and the last two can be examined separately. In this way the problem can be approached as a clustering problem with 30,000 and 20,000 elements in three and two clusters respectively. However, there is a difficulty in ensuring that the mode shapes for a given measurement are in different clusters. This was not taken into account in the initial approaches and the results obtained were verified on this basis and visually.

We could easily reproduce the clustering that we thought was correct for the first three mode shapes.

To do this, we normalised the data and added the first order differences to obtain vectors with 51 elements, then the KMedoids algorithm was used for clustering. The mode shapes belonging to the same mode can be opposite to each other, so a metric was needed that could cope with this. We therefore used the similarity metric known in the field as the MAC value, which is the square of the cosine similarity metric, i.e.:

$$\mathrm{MAC}(x, y) = \frac{(x \cdot y)^2}{\|x\|^2 \|y\|^2}.$$

Using the above method, the first three clusters were easily identified: the clusters all had the same size of 10,000 elements and no visual anomalies were found. It is important to note that they are so different from each other that it was sufficient to train the model on only 2-3% of the data (using random sampling), resulting in especially fast training and prediction.

For the last two mode shapes, several problems were encountered. One is that, unlike the first three, the mode shapes are much more varied, making the visual control less convincing. The bigger problem, however, is that we cannot be sure that the measured mode shapes are indeed from these two (fifth and sixth) and not possibly belong to additional (unmeasured) mode shapes.

For the last two mode shapes, the previous approach resulted in clustering where the cluster sizes were nearly equal, 10,032 and 9,968 in the example shown in Figure 4. Running with a few different hyperparameter values also gave similar results, about 30-40 deviations from the 10,000 required for clusters of equal size.

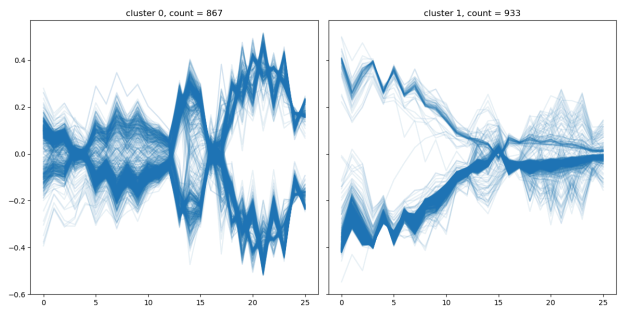

Figure 3: Clustering of a random sample of 1,800 of the last two mode shapes. These are 9% of all records.

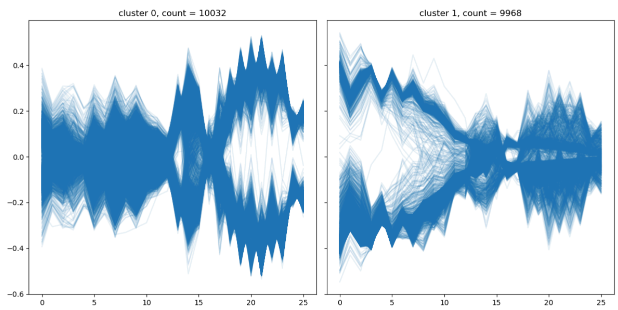

Figure 4: Clustering of the last two mode shapes based on the built model shown in Figure 3.

However, these are only observations on the size of clusters for the time being. Comparing the results with the original dataset, we found that the mode shapes belonging to a record were indeed all in different clusters for the first three mode shapes. However, for the last two mode shapes, we also found 30-40 incorrectly clustered records, almost as many as the deviation of the cluster sizes from 10,000. It is promising that, apart from these few exceptions, the model is clustering plausibly. The reasons for these outliers are not clear and their understanding is the subject of further research.

Conclusions and further research

One of the main issues in improving the model is how to use the parameters in clustering. We expect similar parameter values to be associated with similar mode shapes, i.e. for a given mode shape, even a small change in the parameters will result in a small change. We expect a similar property for eigenfrequencies. Previous attempts, including the one presented in the previous section, have not exploited these properties.

Two main approaches to the former have emerged so far, but no real results have been achieved. The first approach is based on the intuition that we should divide the parameter space into smaller chunks, within which we can hopefully handle the task better, and then build a unified model based on these. The main issue here is the merging of the model.

The second approach is based on a modification of the similarity metric to consider mode shapes with similar parameters as more distant. A similar regularisation is also desired for eigenfrequencies. We have already experimented with this approach, but the question is what metrics can be used to evaluate a solution in this unsupervised learning task, and what weights and regularisation functions should be used to achieve this.

Bibliography

[1] B. Kurent, N. Friedman, W. K. Ao, and B. Brank, “Bayesian updating of tall timber building model using modal data”, *Engineering Structures*, vol. 266, p. 114570, 2022, doi: https://doi.org/10.1016/j.engstruct.2022.114570.