Improving efficient seismic collapse intensity measures using machine learning

Pablo Heresi & Hector Davalos

STANFORD UNIVERSITY

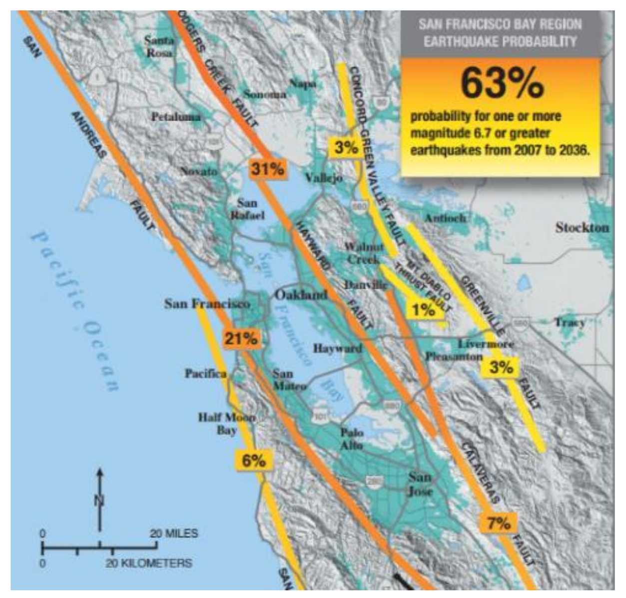
CS 229
Fall 2016

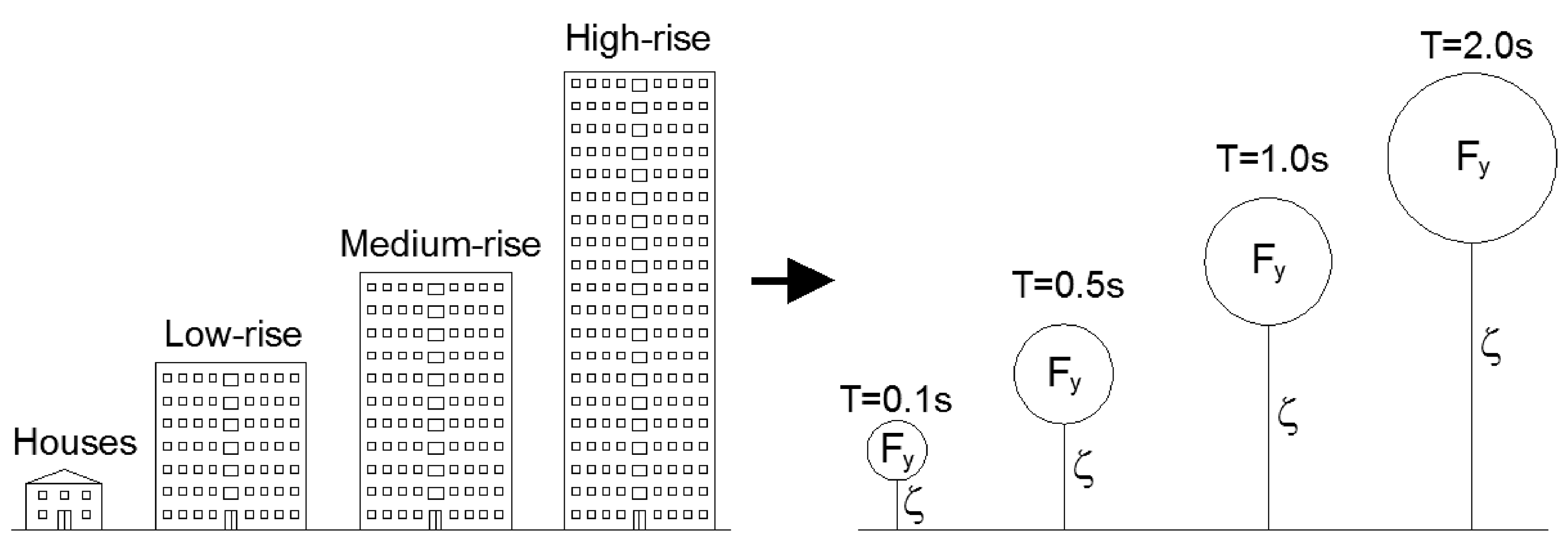
Predicting:

In Earthquake Engineering, a key objective is to estimate the probability of collapse (P(C)) due to ground motions [1]. These predictions are made using scalar parameters named intensity measures (IMs) mainly derived from the records. Then, the main objective of this project is to improve current predictions by finding a set of features for predicting seismic collapse of a set of structures by using SVM and logistic regression. As the P(C) of well-designed structures is relatively low (<10% at the 2,475yrs return period earthquake), we will be dealing with models that should be able to capture relatively rare events.

Data:

- 1680 acceleration time histories recorded in events that are representative of the seismic hazard in the Bay Area: M_w>5, R_{jb} < 100km.
- We will simplify the building stock the city of San Francisco using 4 different single-degree-of-freedom systems having 5% damping (ζ).
- Lateral strengths (F_y) defined accordingly to achieve collapse probabilities within code recommendations.





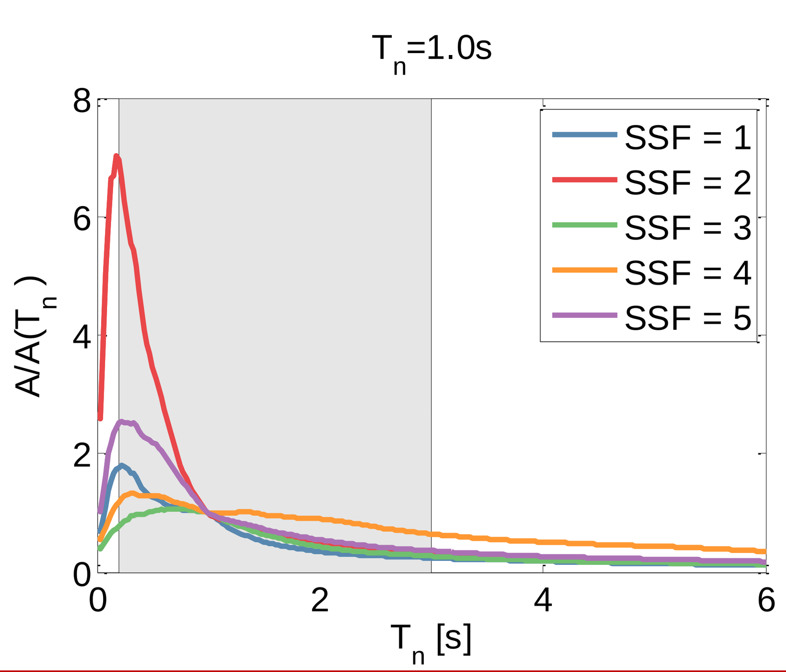
Attributes:

A set of 10 different parameters ($x \in \mathbb{R}^{10}$) computed from record characteristics:

1) Earthquake Magnitude (M_w)
2) Rupture-to-site distance (R_{jb})
3) Peak ground acceleration (*PGA*)
4) Peak ground velocity (PGV)
5) Spectral ordinate (*Sa(T)*)
6) Average Acceleration (*SaAvg*) [2]
7) Incremental velocity (*IV*)
8) Filtered IV (*FIV*)
9) Spectral shape factor (*SSF*)
10) Significant duration (*SD*)

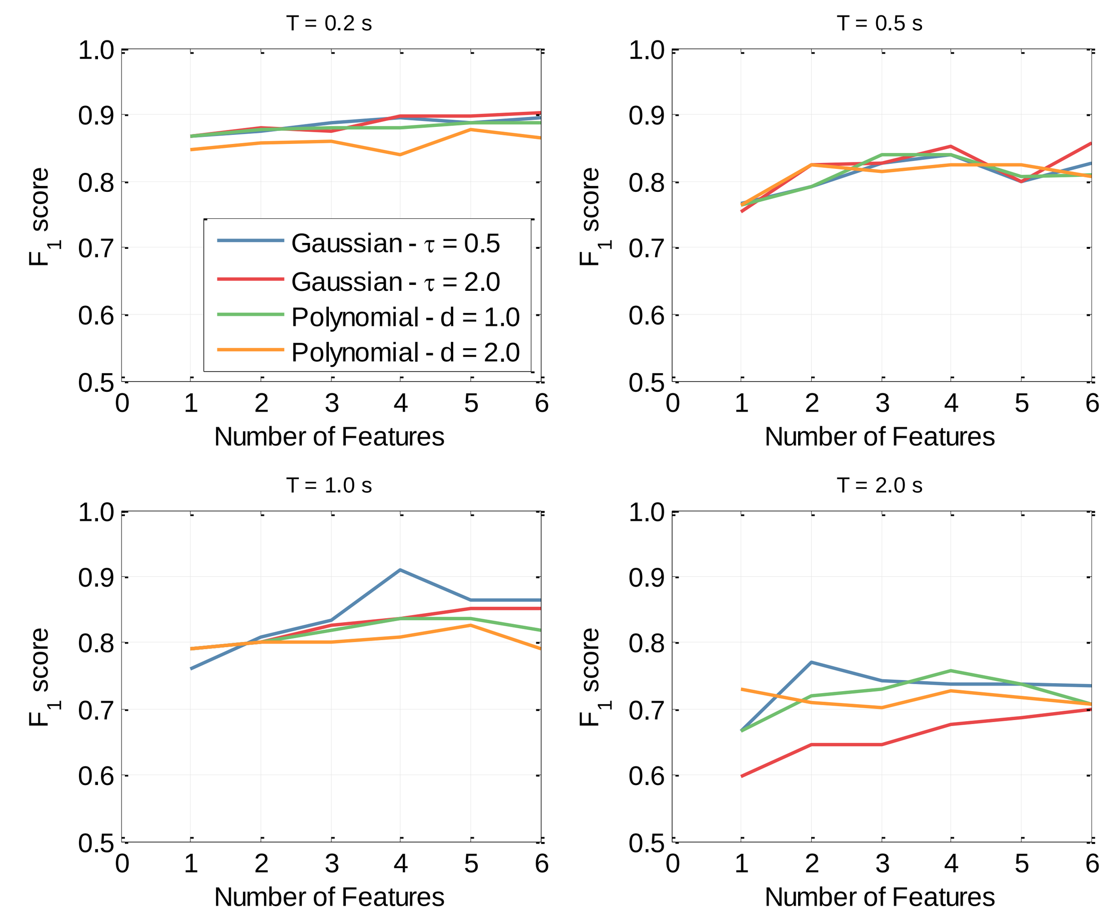
K-means clustering of normalized spectral shapes (SSF)

Used to separate all records into several spectral shape types as it well know that spectral shape influences lateral displacement demands. The range used for clustering depends on T(0.2T–3T) as previous research has suggested [2].



Results:

SVM Feature selection 70/30 Hold-out cross validation



Confusion Matrices

Optimal SVM with 3 parameters

Gaussian kernel, τ = 0.5
log(SaAvg), M_w, R_{jb}

T = 0.2 s		Predicted -1	Predicted 1
True	-1	1492	15
True	+1	30	143

Precision = 0.91
Recall = 0.83
F1-score = 0.86

1st degree polynomial kernel
Sa(T), SD, FIV

T = 0.5 s		Predicted -1	Predicted 1
True	-1	1553	28
True	+1	32	67

Precision = 0.71
Recall = 0.68
F1-score = 0.69

Gaussian kernel, τ = 0.5
SaAvg, SD, IV

T = 1.0 s		Predicted -1	Predicted 1
True	-1	1552	14
True	+1	30	84

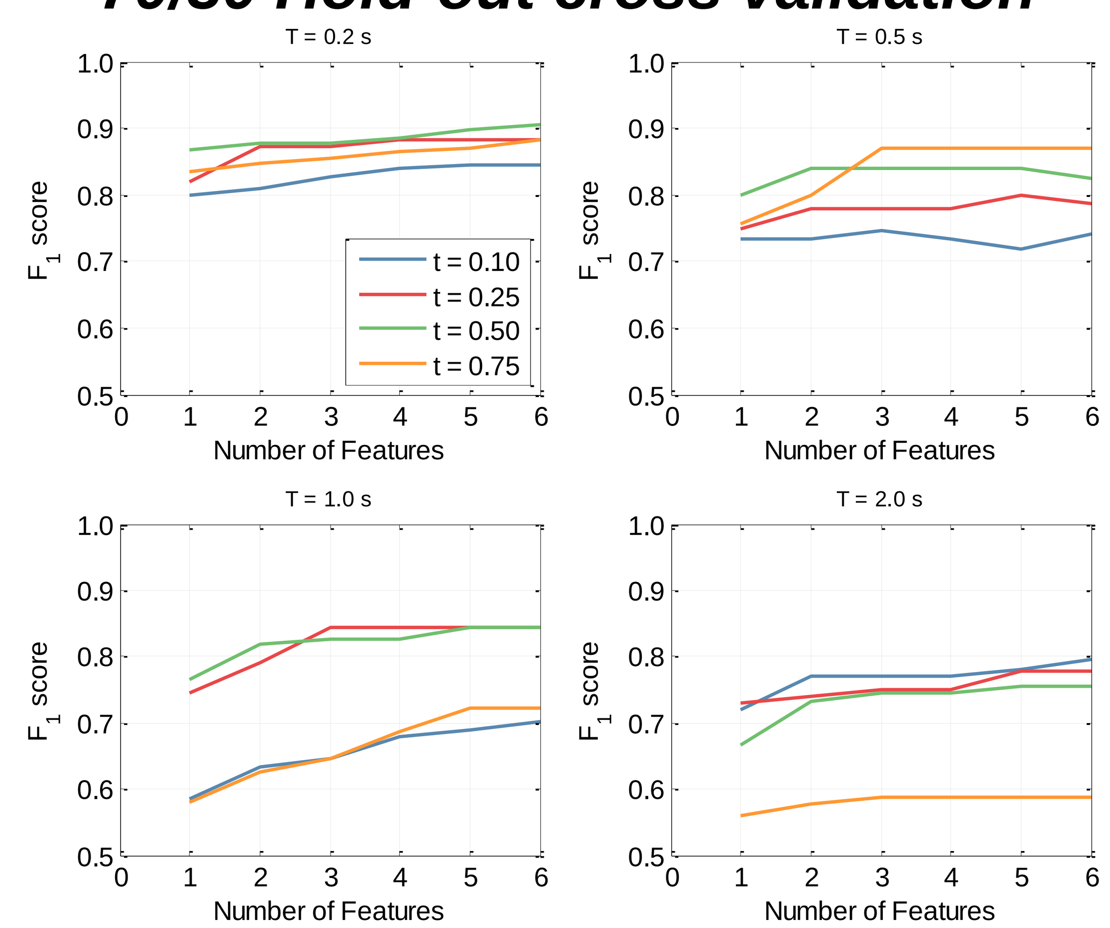
Precision = 0.86
Recall = 0.74
F1-score = 0.79

Gaussian kernel, τ = 0.5
log(SaAvg), Sa(T), PGV

T = 2.0 s		Predicted -1	Predicted 1
True	-1	1452	47
True	+1	45	136

Precision = 0.74
Recall = 0.75
F1-score = 0.75

Logistic Regression Feature selection 70/30 Hold-out cross validation



Logistic Regression with 3 parameters

Our Recommendation

threshold = 0.5
log(SaAvg), PGA, Sa(T)

T = 0.2 s		Predicted -1	Predicted 1
True	-1	1489	18
True	+1	37	136

Precision = 0.89
Recall = 0.79
F1-score = 0.83

threshold = 0.75
SaAvg, SSF, SD

T = 0.5 s		Predicted -1	Predicted 1
True	-1	1579	2
True	+1	45	54

Precision = 0.96
Recall = 0.55
F1-score = 0.69

threshold = 0.25
SaAvg, M_w, SD

T = 1.0 s		Predicted -1	Predicted 1
True	-1	1536	30
True	+1	20	94

Precision = 0.76
Recall = 0.82
F1-score = 0.79

threshold = 0.1
SaAvg, R_{jb}, M_w

T = 2.0 s		Predicted -1	Predicted 1
True	-1	1377	122
True	+1	7	174

Precision = 0.59
Recall = 0.96
F1-score = 0.73

Discussion:

In general we see that *SaAvg* is a very good predictor of collapse but the 2nd and 3rd features vary depending on the type of structures.
Results from the logistic regression were almost as good as the ones obtained using SVM, however, due to the time it takes for the SVM to learn from the data, we choose logistic regression as the optimal method.
For earthquake engineering purposes, it is practically unfeasible to use more than three parameters as IMs because as part of the probabilistic seismic demand analysis, one needs to estimate the probability of observing different levels of the parameter used to characterize collapse. Therefore, we recommend a three-parameter vector that gives the best results for each typology as we have to develop procedures to estimate the annual rate of exceeding different levels of these new IM vectors. This information would be represented using a joint hazard curve [3]. Then, by integrating the probability of collapsing at a certain IM level with the probability of experiencing that IM level we can compute the mean annual frequency of collapse.

Future work:

- Develop a ground motion prediction equation of the joint distribution of these new IM vectors.
- Make our predictions structure-type specific: Masonry, Concrete, Steel. This will imply modeling the structural response using different force-deformation behaviors instead of the simplified one used in this study.
- ***Different and promising idea: Apply Markov decision processes in active control of structures during earthquakes.***

References:

[1] C. A. Cornell, and H. Krawinkler, “Progress and challenges in seismic performance assessment”. *PEER Center News*, vol. 3, pp. 1-3, 2000.
[2] L. Eads, and E. Miranda. Average spectral acceleration as an intensity measure for collapse risk assessment. Earthquake Eng. And Struct. Dyn., vol. 44, pp. 2057-2073, 2015.
[3] P. Bazzurro, and C. A. Cornell. “Vector-valued probabilistic seismic hazard analysis”. *Proc. of the 7th US national conference on earthquake eng.*, pp. 21-25, 2002.