Unsupervised Techniques for Feature Extraction in Seismic Noise Data

Dylan Snover | dsnover@ucsd.edu
Derek Lam | ddl004@ucsd.edu
Brian Whiteaker | bwhiteak@ucsd.edu
Anish Chivukula | anchivuk@ucsd.edu

Background

Seismic data was recorded over the course of 4 weeks on a spatially dense array consists of ~1108 vertical component geophones on the San Jacinto Fault Zone. Using a CNN, the dataset was separated into three classes: Earthquake Data, Random Noise, and Non-Random Noise[1](NRN). Our group worked exclusively with the Non-Random Noise, ~59-74% of the original dataset[2].

- Goal: utilize unsupervised techniques to extract features from NRN and use these features to cluster NRN into separable classes.
- Motivation: Differentiating NRN may lead to source identification.

Data

- The dataset was provided in the form of spectrograms generated from 4-second wavelets sampled at 500Hz.
- Dataset consists of 520,941 spectrograms.
  - 50,000 spectrograms were isolated from dataset to be used as test data.
  - Remaining 470,942 spectrograms were split into 80% training data (376,753 samples) and 20% validation data (94,188 samples)

Features

Each spectrogram in the dataset has a height of 251 pixels and a width of 41 pixels, totaling 10,291 features.

- Each feature of the spectrogram represents the amplitude of a certain frequency at a given point in time.
- Obtaining a smaller latent space through feature extraction methods will allow for more effective results when performing clustering techniques.

Models

Feature Extraction

- Convolutional Autoencoder (CAE): The CAE employed 3-layer encoder and decoder modules. For each encoder layer we had sequence of 2D conv, batch normalization, ELU, and maxpool. This was mirrored in reverse by the decoder. The ELU non-linearity stands for exponential linear unit. For a loss function we utilized MSE and gradient descent was optimized with RMSprop. The best learning rate was found to be lr=0.01. Weight decay underperformed.

- Non-Negative Matrix Factorization (NMF)[3]: Given a data matrix **X**, NMF factorizes this matrix into a basis matrix **W** and reduced latent space **H**, where all three matrices contain non-negative values and such that **X = WH**. **W** and **H** were initialized with NNDSVD.

Clustering

- K-Means Clustering: Each sample is assigned to the cluster *i* whose mean has the lowest squared Euclidean distance:

$$i^*(x) = \underset{1\leq i\leq k}{arg\,min}||x-\mu_i||^2$$

  The mean of each cluster is recomputed after each iteration.

- Hierarchical Agglomerative Clustering (HAC): At first, each point is assigned its own cluster. Each cluster merges with other clusters based on the specified distance metric and linkage method.

- t-distributed Stochastic Neighbor Embedding (t-SNE)[4]: t-SNE is able to map a high-dimensional space into a 2-dimensional space by attempting to minimize Kullback-Leibler divergence between the two spaces, helpful for visualization of our data.

$$cost = \Sigma_i KL(P_i||Q_i) = \Sigma_i \Sigma_j p_{j|i} log \frac{p_{j|i}}{q_{j|i}}$$

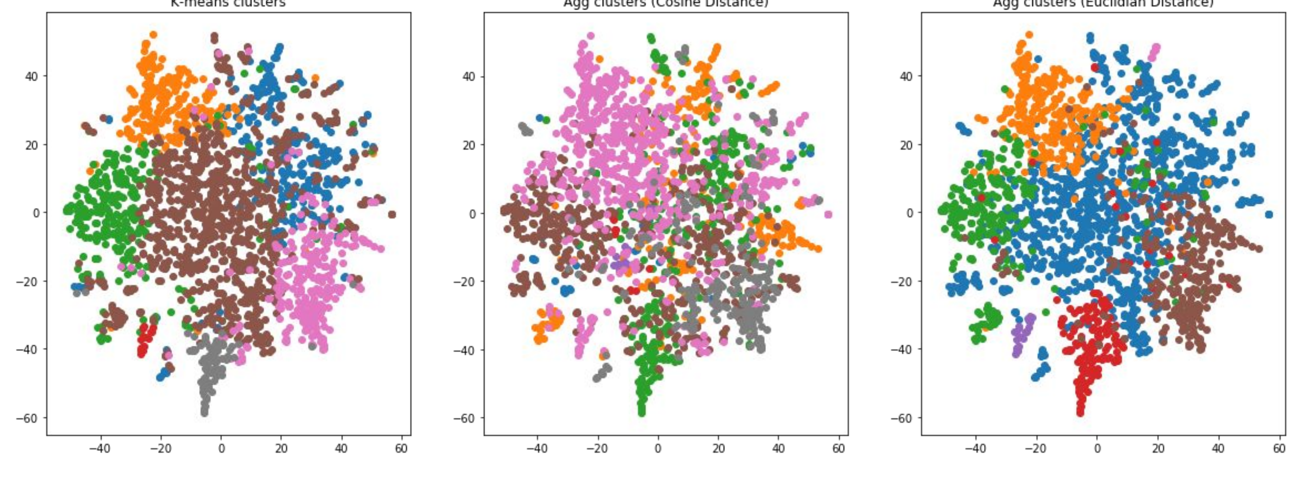

Figure: t-SNE visualization of clustering after NMF feature extraction, using K-means, HAC with cosine metric, HAC with Euclidean metric

Results

Reconstruction Results

Reconstruction Errors on Train, Validation, and Test Sets (MSE)			
Model	Training Error	Validation Error	Test Error
CAE	6.30e-05	2.53e-04	9.16e-03
NMF	2.86e-05	2.87e-05	2.85e-05

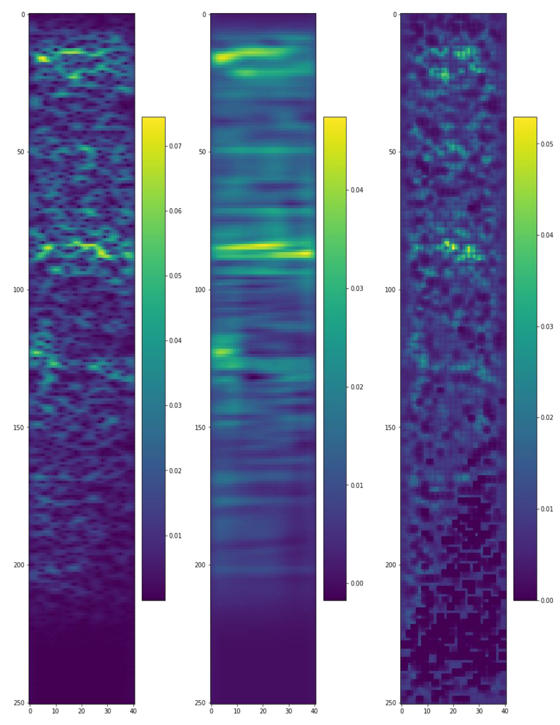

Figure: Original Spectrogram from Test Set (left), Reconstruction with NMF using a latent space with dimension 250 (middle), Reconstruction with the CAE using a latent space with dimension 288 (right)

Clustering Results

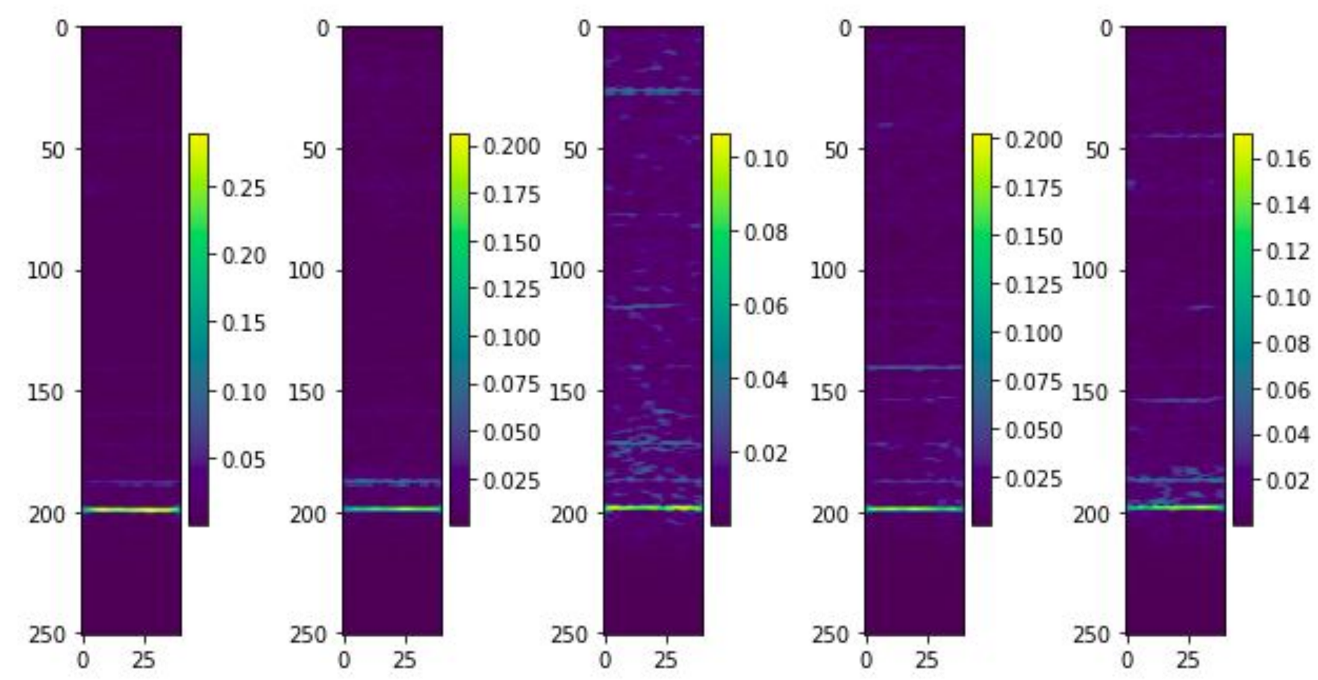

Figure: Example of a subjectively good cluster using K-Means on reduced feature space obtained from NMF

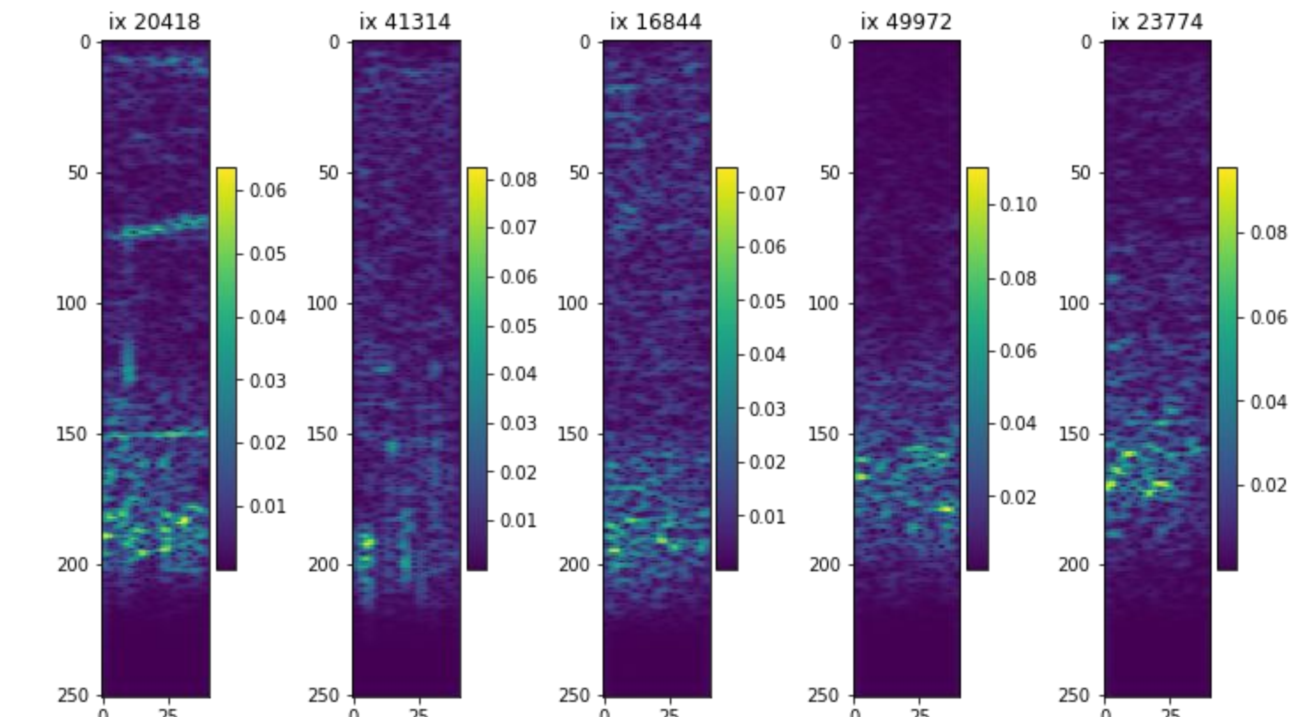

Figure: Example of a subjectively good cluster using Hierarchical Agglomerative Clustering on a reduced feature space obtained from CAE

Discussion

The overall goal was (1) to reduce the dimensionality of the data through feature extraction (2) use unsupervised learning to cluster the spectrograms.

- NMF required a large amount of computation power for training due to the size of the data, but provided a robust model for feature extraction.
- The CAE required exploration of architecture, loss function, optimizer, non-linearity, depth, and filter size.
- Although the examples above showcase subjectively good clusters, there are many outliers within these clusters. More work will need to be done.

The CAE had issues with abstraction and depth of network which required balance. A future approach may involve segmenting the image into thirds and applying the CAE separately, concatenating the outputs at the end. The aim would be to limit the CAE from abstracting out location information.

References

1. Ben-Zion, Yehuda & Vernon, Frank & Ozakin, Yaman & Zigone, Dimitri & E. Ross, Zachary & Meng, Haoran & White, Malcolm & Reyes, Juan & Hollis, Dan & Barklage, Mitchell. (2015). Basic data features and results from a spatially dense seismic array on the San Jacinto fault zone. Geophysical Journal International. 202. 1-11. 10.1093/gji/ggv142.
2. Johnson, C. W. (2019). Machine Learning with CNN in Seismology [PowerPoint slides]. Retrieved from http://noiselab.ucsd.edu/ECE228/slides/lecture9.pdf
3. I. Buciu, "Non-Negative Matrix Factorization a New Tool for Feature Extraction: Theory and Applications", Int'l J. Computers Comm. and Control, vol. 3, pp. 67-74, 2008.
4. van der Maaten, L. & Hinton, G. E. Visualizing data using t-SNE. J. Mach. Learn.Research 9, 2579–2605 (2008).

UC San Diego
Jacobs School of Engineering