Spaceapps Presentation Sparknotes

Hilbert Huang Transform

* The Hilbert Huang transform is an emerging technology which has only come into the signal processing scene recently.
* It allows one to take a signal and decompose it into functions which contain information about the different things happening within the signal, such as instantaneous frequency and phase.
* These modes are given in a sequence, one is the primary mode, secondary, and so on. The given data typically has 5 or 6 unique modes, but only the first few contain important information.
* The difference between this and a Fourier transform is that it allows you to retrieve aperiodic information from a signal, which is crucial when we are dealing with signals that are not wavelike
* We use it here in order to denoise our data and develop correlations between different locations data sets’ by taking the correlation coefficient of two columns secondary modes.

Outlier detection system:

* Big picture: the algorithm scores pairs (HH-DD, Location) based on whether other locations experience a similar disturbance at the same time. This lets us track down local (rather than global) interferences such as storms, power outages, and other technical difficulties.
* Factors in the scoring method:
* Correlation function between the two signals (Filtered by a HHT)
  + After taking the Hilbert Huang transform and finding the intrinsic modes, we find the correlation between each of the columns’ secondary mode. The secondary mode contains much of the information about peaks and valleys in our data, but is filtered of noise – this allows us to remove noisy signals from the scoring data which makes the result extremely clean in comparison to just using the STFT and distance model.
* Maximum-norm between the moving (STFT) Fourier transform of two signals
  + This means that we are taking the fourier transform of a small interval (15 samples wide) for each column. If our column of interest is column x, then we sum together the absolute difference between column x’s transform and the other columns. This gives a direct measure of difference between the signals.
* Physical distance (along Earth’s surface) between points at which signal is measured
  + If two measurements are taken close to eachother and both have the same extraneous signal, it is likely that the signal is due to some kind of local phenomenon, and is probably not a technical difficulty. Thus we multiply the score by the distance.

Results

* On these two slides, flip back and forth to compare between the plots
* The heatmap is the normalized score, but the unnormalized score is what is saved in the text file.
* You can visually see a few of the outliers, which gives credence to the model
* The data and script is on our github (see the last slide for a link)

Interface concept

* The project can be extended by fine tuning the model and by building a nice GUI for user friendliness
* We did not finish the interface

Applications

* The algorithm we’ve put together is easily tunable, with more fine tuning the results would improve even more
* It can be used to reliably detect anomalous data in most sets, regardless of the data type or quantity, most notably non-linear and non-stationary data
* The Hilbert-Huang Transform has a lot of potential as an emerging technology in signal processing, which is used by any scientists working with large time-series datasets
* This algorithm won’t change the world, but rather, it will make changing the world easier

If anyone asks, there is a supplementary slide too with the score equation.