Dendrogram Clustering of fMRI Data to Improve SNR while Maximizing Information Content

Rafael Bravo
Computer Science Masters Program
rbravo@ucdavis.edu

Abstract:

  This project concerns the application of clustering to fMRI time-series data. By clustering voxels, the noise in the data can be mitigated, as averaging over multiple correlated voxels will increase the signal to noise ratio significantly over examining individual voxels. Here a clustering algorithm using dendrogram clustering is proposed, as this method optimally combines similar voxels in a stepwise fashion. To account for localization of function in the brain, only the 6 voxels directly surrounding a given one are considered for clustering. To assess the information loss at each step, the Entropy of the system is calculated after every iteration. Two different clustering parameters are here considered, one being correlation coefficient, the other being mutual information. The results here point to correlation coefficient as a superior method for conserving system entropy.

Introduction:

  Current fMRI data analysis is typically performed using one of two methods; Either voxels are correlated with time based events on an individual level, and the researcher looks for clusters of voxels that all correlate with the experiment, or specific regions of interest (ROIs) are defined a-priori, and only these voxels are assessed. In the first case, clusters are formed based on how well voxels correlate with the experiment and and just the experiment, without first assessing structural significance.
  ROI selection takes the opposite approach, the specification of a specific region before seeing results forces analysis to only take place in specific areas and prevents simply fishing for correlations, however this method has a hard time taking the individual data into account. Since all brains are different, and regions of interest can vary significantly in size and location, it can be difficult even with alignment to a template space to choose a voxel region that works for all subjects.
  Clustering the brain into discrete regions before taking the data into account takes the intermediate path between these two methods. On the one hand, after clusters are formed, the researcher can still fish them for correlations with experiment events, however the "pool" to fish from is significantly smaller. More importantly, these clusters were created by clustering similar voxels, and is based purely on correlations within the brain data, and not in a manner specific to the experiment events. In this way it is similar to the ROI selection a priori, where regions are specified before looking specifically at experiment correlations, but as an added bonus these regions are guaranteed to be aligned with the data. While it is not immediately apparent what clusters correspond to what brain regions, as with the a priori ROI selection, it is hoped that by clustering in a manner that conserves as much information as possible, that regions that behave contiguously will naturally emerge out of the data and be recognizable in terms of location and function. This is the motivation for

this project.

The clustering algorithm here proposed uses dendrogram clustering, which starts with the brain separated into individual voxels, and then clusters the two most correlated, spatially proximate voxels. this process is repeated, and can be iterated until the entire brain is one big cluster. Using this method, there are many important questions to consider. This project addresses three of these questions specifically: First, what is a good time to stop clustering so that the clusters will be as large as possible without losing too much information? Second, what is a good measure for similarity between voxels to promote the generation of meaningful clusters? Third, should the newly computed average cluster values be used for subsequent steps in clustering or should the original data values be used?

Background:

Functional Magnetic Resonance Imaging (fMRI) is a very promising method for gathering data on the entire brain volume during an experiment. fMRI involves iteratively scanning the brain with a strong electromagnet and detecting the magnetic response to perturbation, yielding a time-series of values. When nerve cells fire and exhaust resources, oxygenated blood is pumped to those cells to replenish them. fMRI works by detecting this hemodynamic response, and measuring changes in blood flow over an approximately 6 second period directly following activation.

fMRI scans the brain in slices, and usually for analysis purposes these slices are interpolated together into single time slices.Complete scans are taken approximately every two seconds. to standardize the positions of the voxels, scans are also typically aligned to correct for motion, as well as warped to a standard space to help account for differences across individuals.

As can be seen based on these standard preprocessing steps, as well as the indirect nature of the hemodynamic response, fMRI study is far from an exact science. individual voxels are usually quite noisy as well. the typical method for dealing with this low signal to noise ratio is blurring the data via a Gaussian function, which can essentially average out noise, but at the cost of signal definition.

The current approach follows these preprocessing steps before the data is clustered, with the exception of blurring, as the averaging of values within clusters should serve a similar function.

Methods:

The methods used to analyze the system can be classified into three types. The first set of methods involve the preprocessing of the fMRI data, described in the background. The second set of methods involve the clustering algorithm itself, these are the dendrogram clustering algorithm and the similarity metric used to generate clusters. The third set of methods involve analysis of how the system changes with more cluster formation, the metrics used here are the entropy of the system as well as the size of the largest cluster.

The data preprocessing for this project was done using the fMRI analysis software AFNI. A single subject resting state brain scan taken over around 440 seconds (a total of 220 volume samples) was used as the data source for

this project after assessing the data to ensure that there was not excessive motion or other data corrupting factors. The brain size was also scaled down by a factor of 2 (by averaging voxel values in 2x2x2 squares) to increase the speed of computation. This lowered the total number of voxels per time sample down from 44471 to 5263.

   The clustering algorithm was written in c, interfaced with a framework for viewing and preparing the data written in python. First the similarity measures between all touching voxels were calculated and put into a list. the voxel pair with the highest similarity rating is chosen from this list, and the two voxels that share this similarity were grouped together into a new cluster. Any voxels that were already part of a cluster with either of these two voxels are also merged with the new cluster. The algorithm starts with all voxels as separate clusters, and ends with the entire brain merged into one cluster. Only choosing the nearest neighbor was made a constraint to force contiguous group formation as well as because it made the clustering algorithm run faster as fewer similarity measures needed to be calculated.

   Two similarity measures were compared in this project. One of these was the mutual correlation coefficient between voxels, the second was their mutual information. the number of bins used to calculate mutual information was also varied, looking at 2, 10, and 100 bins, bound evenly between the minimum and maximum value over the volume.

   To measure the information loss occurring with each step, the entropy of the system was measured with each new iteration of the clustering algorithm. To calculate the entropy over the entire experiment, entropy values were calculated across all voxels in the brain volume at a given time point, then these entropies for all times were averaged. Similarly to the mutual information clustering algorithm, the number of bins used to calculate the entropy was also varied between 2, 10, 100, and 1000 bins.

Results:

   There are several interesting findings from this data analysis. The first question is whether to update values for voxels within a cluster when it is formed, or to let voxels keep their original values. This impacts further cluster formation. The argument for updating is that clusters formed by recomputing the averaged correlation with every merge should have an overall higher correlation amongst their voxels. The argument against this is that not tampering with the original correlations may result in clusters that correspond more closely to the original data.

   As can be seen in Figure 1, the entropy of the system declines significantly slower with updating of similarity measures than without. Also, as can be seen in Figure 2, the maximal cluster size stays smaller for a far longer duration when clusters are being updated than when they are not. In fact, when clusters are formed without averaging, it appears as though a single contiguous region will quickly result rather than many small individual clusters, as can be seen by comparing Figures 3 and 5.

   Interestingly, voxels towards the middle of the brain where the white matter is located, are the last regions to form clusters, as is evidenced by Figure 4. This is the same setup as Figure 3, the only difference being that all clusters of size 5 or smaller have been removed. It is expected that clusters would not form there, as fMRI is believed to only be able to detect activation in the gray matter, meaning the white matter regions are

governed by noise.

  We next look at the comparison between using correlation coefficient vs mutual information as the measures of similarity. mutual information was computed using 3 different numbers of bins, 2, 10, and 100. the number of bins that the data was separated into largely affected the mutual information values computed for all voxels. This can be visualized by looking at the changes in entropy and largest cluster size as the number of bins are altered. These data are presented in Figures 6 and 7. in all cases, it is clear that mutual information with only 2 bins, or a binary sorting of values being above or below, does a comparatively poor job of forming useful clusters. 10 and 100 bins do a better job, but Correlation Coefficient is clearly the best at both conserving mutual information and keeping the maximum cluster size low. For this reason, the use of Correlation Coefficient as the similarity metric is recommended.

  Finally, using the preferred settings determined via the previous several tests, the sufficient number of bins needed to test the entropy of the system is assessed. Entropy graphs before this point have been computed using 100 bins. 2, 10, 100, and 1000 bins for computing entropy are compared. Figure 8 shows the results of graphing the entropy change with these numbers of bins. There are several noticeable differences in how the Entropies change over time for the different systems. in the case of small numbers of bins, the entropy is unstable, fluctuating depending on how the values in the system are changed. For larger numbers of bins, however, there is consistent information loss with each step. The reason for this is that Entropies taken with fewer bins are already losing significant amounts of information simply by coarse graining the resolution of the system. Therefore, simply by chance, the system may attain a state that seems more entropic by changing its value distribution. even if these values are in fact more similar.

Conclusions:

  There are several clear results that this work reveals. The first of these is that Mutual Information does not seem as good a metric for conserving Entropy in the system as correlation coefficient, no matter the number of bins. The second is that averaging cluster values at each step seems to produce better results than not doing so. Finally, a comparatively large number of bins, in the 100s, seems a better measure of the amount of information lost than a lower number, as greater numbers tend to show consistent loss rather than sporadic information gain.

Dynamical System/Future Work:

  The brain is a complex nonlinear system that may initially seem difficult to apply information theory concepts to. The amount of data gathered through fMRI scans is paltry in comparison to the amount of activity that is occurring.

  Because observation of the system is so difficult and abstracted, and still so little is known about the brain's function on a larger level, it is difficult to treat the brain as a Dynamical System with defined equations of motion. However, a potential future direction to take this work could be attempting to form Markov models from the cluster data. clusters are a nice target for Markov modeling because the signal to noise ratio for clusters is low, and the clusters form over regions of similar

behavior, meaning that a model of their aggregate data would do a good job describing their constituent voxels. these methods could be used to find pairs or groups of clusters that would not be found simply through correlation, but are nonetheless related structural-temporally.
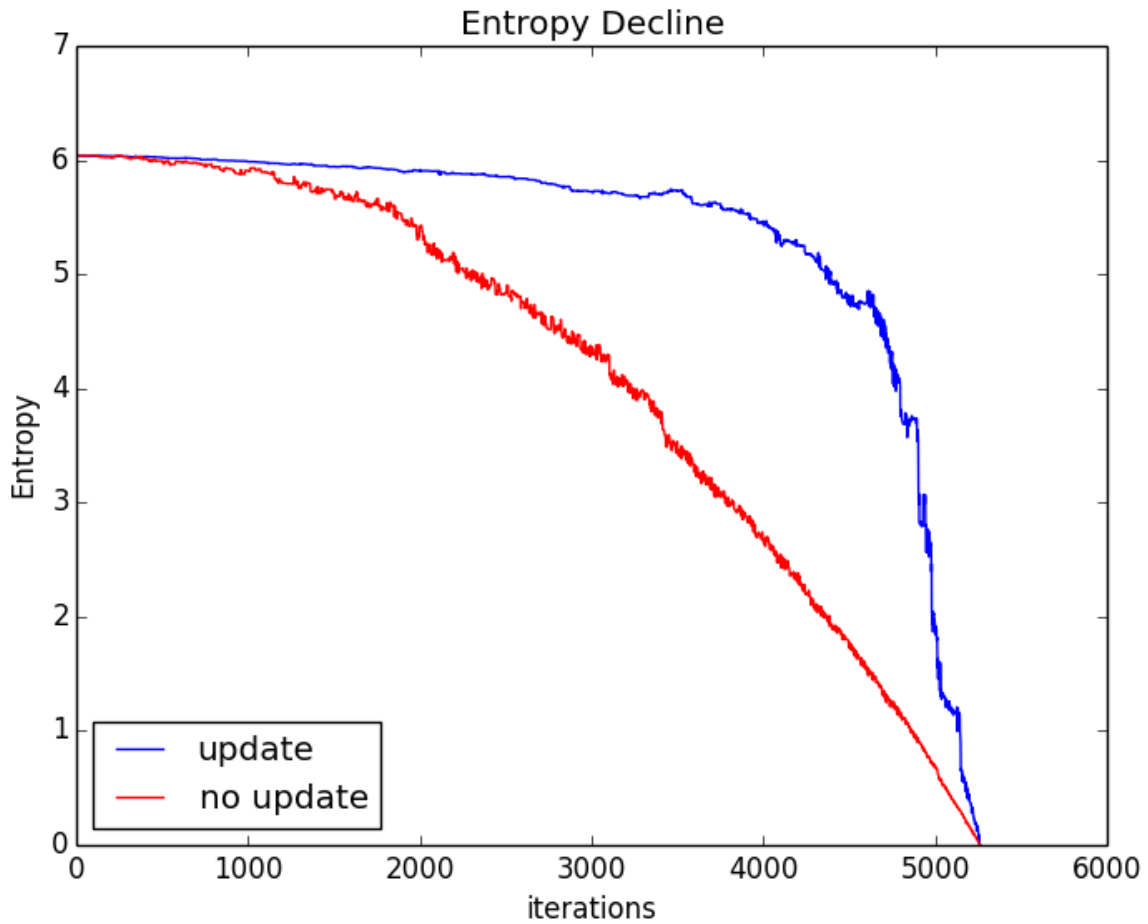Figures

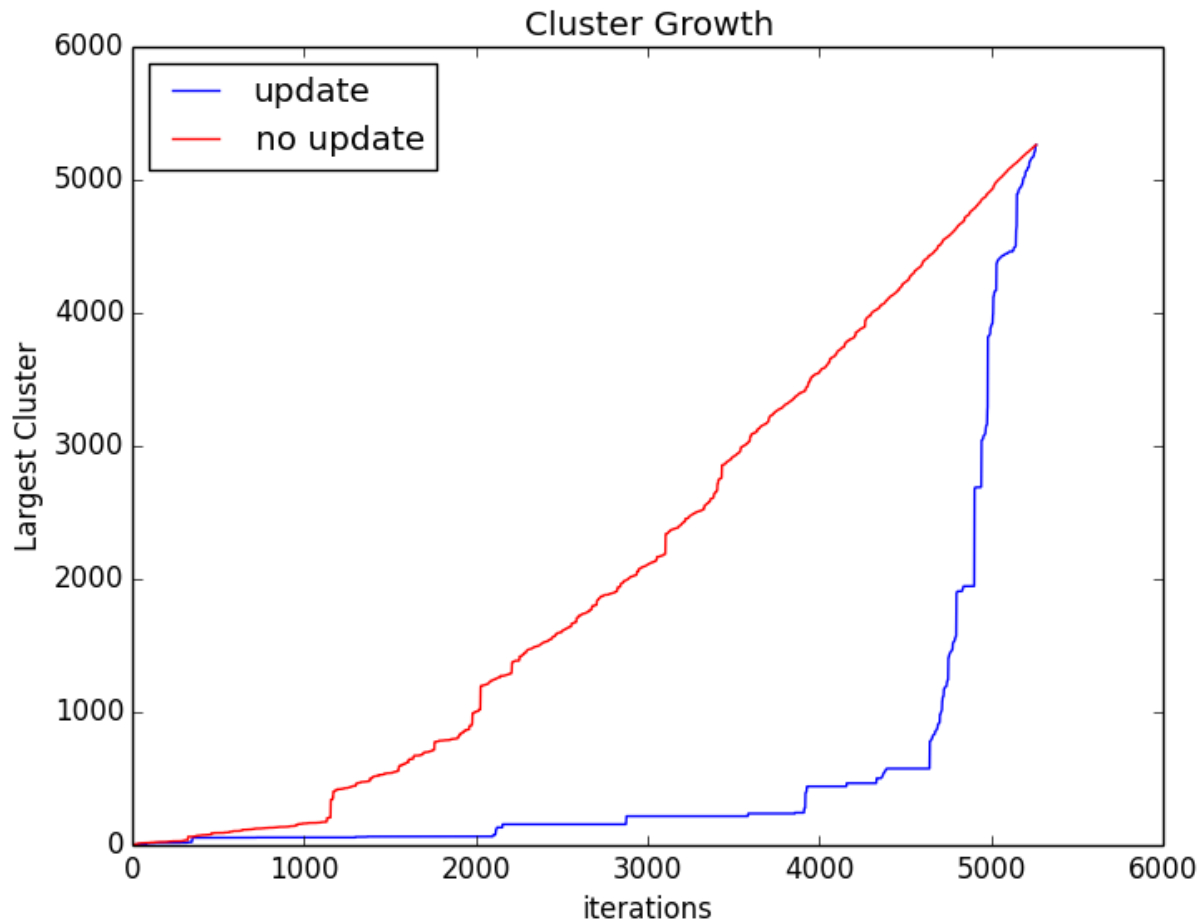Figure 1: Updating cluster values vs not doing so
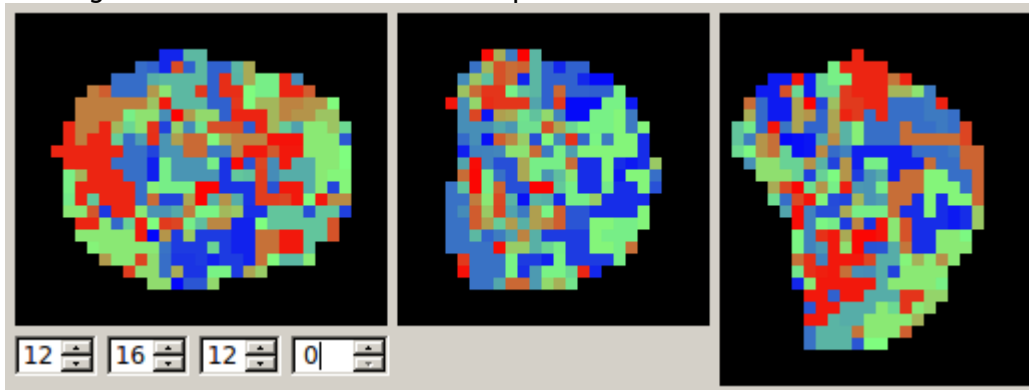
Figure 2: largest cluster size with updates vs without

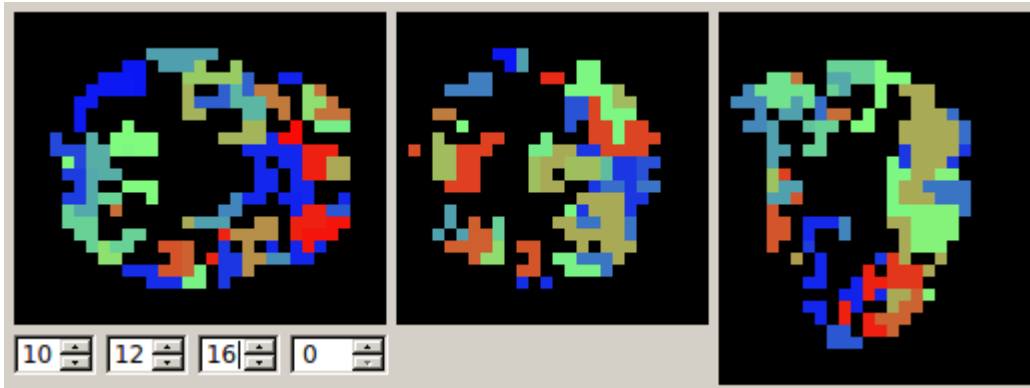Figure 3: sample brain slices after 3800 iterations, using clustering coefficient metric

Figure 4: sample brain slices after 3800 iterations, using clustering coefficient metric, but with clusters of size 5 or less removed.
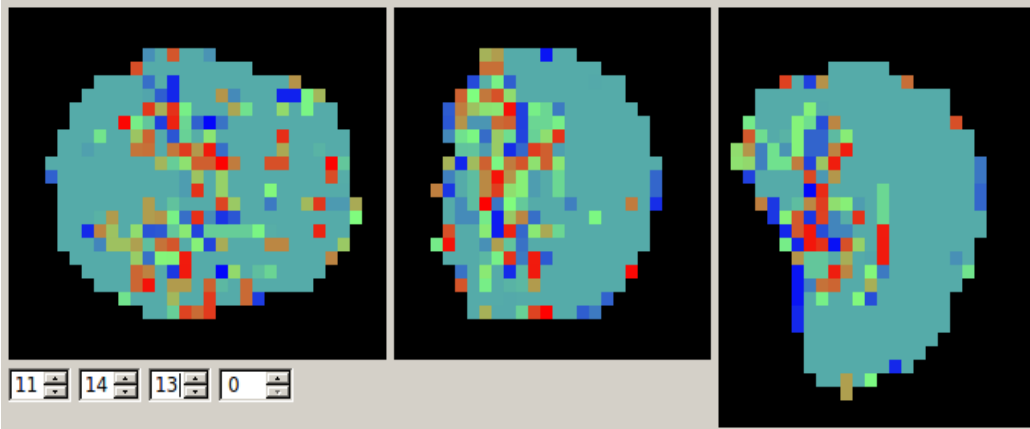

Figure 5: sample brain slices after 3800 iterations, without updating cluster values.
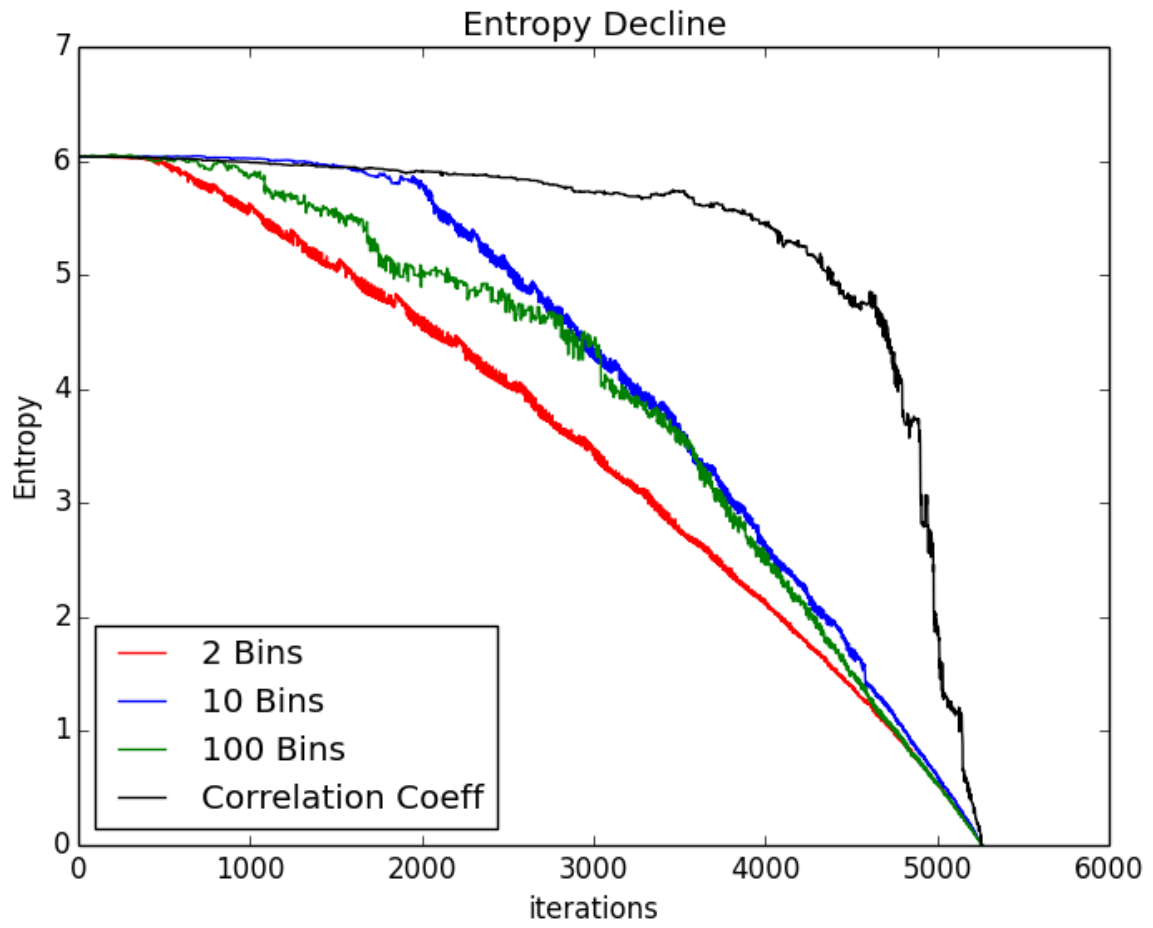
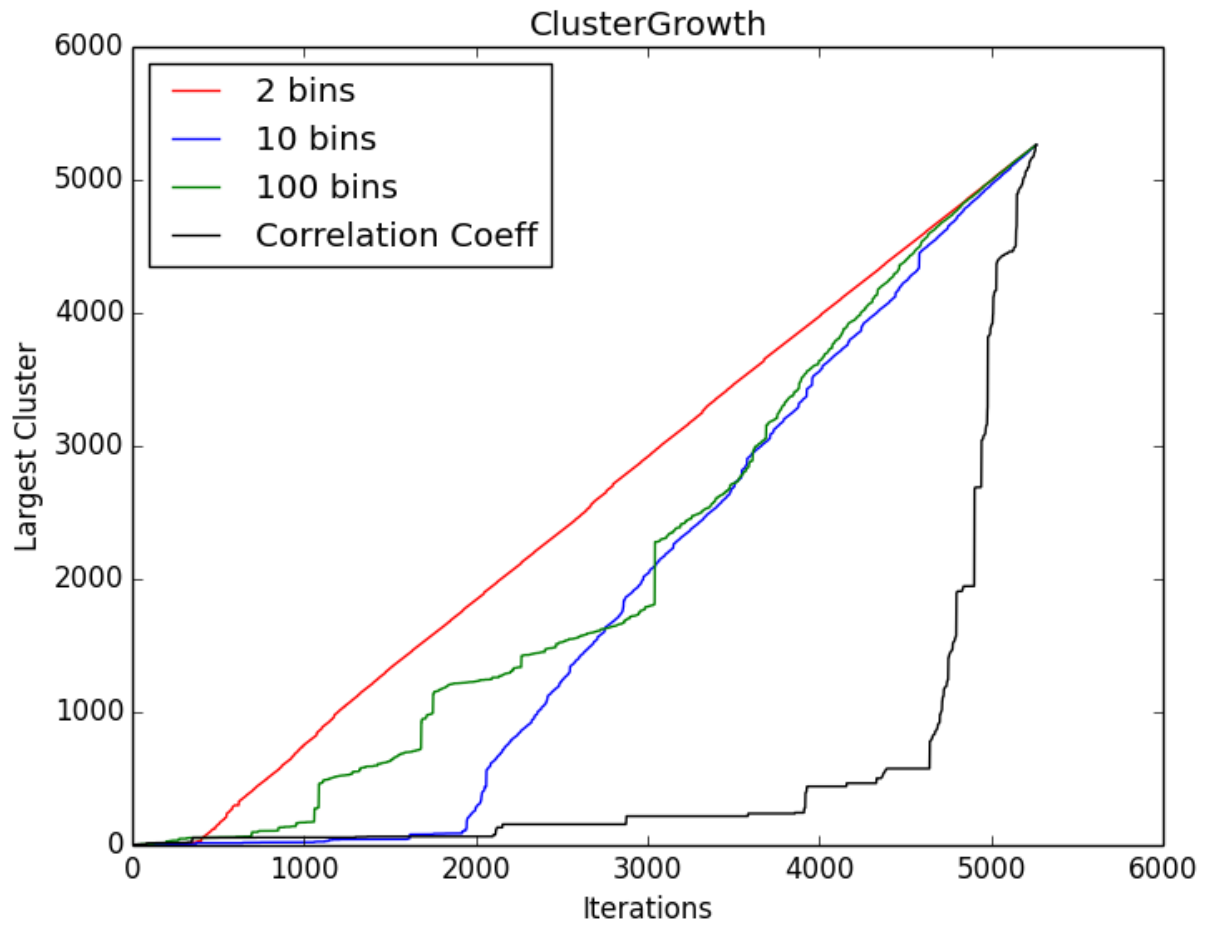Figure 6: comparing entropy change with different metrics

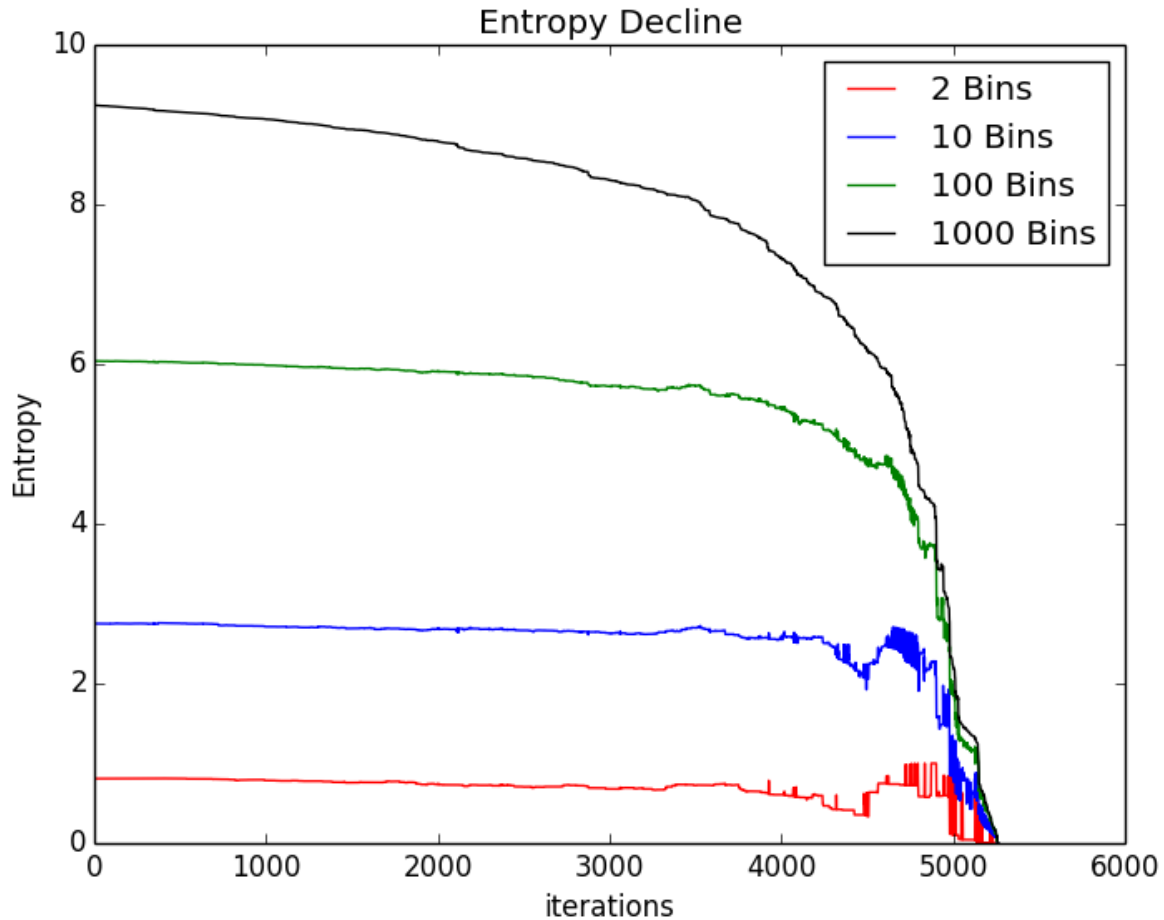Figure 7: comparing change in largest cluster size using different metrics

Figure 8: comparing entropy decline using different numbers of bins to measure entropy

Bibliography

authors, "Paper Title", Journal, vol, pp. #pageNumbers#, date.
Venkatraraman, Van Dijk, Bukner, Golland, et al. "Exploring Functional Connectivity in fMIR Via Clustering" IEEE International Conference, pp. 441-444, 2009.

R. Heller,D. Stanley, D. Yekutieli et al., "Cluster-based analysis of fMRI data", NeuroImage vol 33 pp. 599-608, 2006.

L. Sanberry, R. Nandy, D. Cordes, "Cluster Analysis of fMRI Data Using Dendrogram Sharpening" Human Brain Mapping 20:201-219, 2003.

N. Tavakoli, "Entropy and Image Compression" Journal of Visual Communication and Image Representation vol 4, 271-278, 1993.