

Classifying Brain States Using Causal Brain Networks

Kamal Kamalaldin

I. DEFINITION

A. Project Overview

This project explores the possibility applications of unsupervised learning on classifying different brain states in a rat's brain. The topic is in the neuroscience field, and the motivation behind this project is to help detect epilepsy in the human brain. Although there have been approaches to solving this problem by applying neural networks to EEG signals, and a recent competition has been launched on Kaggle that deals with epilepsy prediction, we take a different approach to this problem by relying on a causality model that is not often used in the field of machine learning. This causality model, Sugihara Causality, is used as a means to establish a measure of brain communication patterns. Using this causality model, we were able to build a series of graphs each of which (theoretically) represents a snapshot of brain communication. Since these graphs are highly dimensional, we use PCA to attempt to reduce its complexity before clustering. we then analyze different clustering possibilities and finally observe if clustering gives an intelligible insight into classifying epileptic and non epileptic brain states.

The dataset used in this project has been collected privately by Dr. Zsolt Borhegyi at MTA-TKI , MTA-ELTE-NAP B-Opto-Neuropharmacology Group in Hungary. The original data consisted of EEG recordings of a rat's brain. The processed (graph series) version of the data can be downloaded from <https://drive.google.com/file/d/0BzyCB-i-aKDWU2pBYkVoOV81S3M/view?usp=sharing>. The research conducted in this paper was also funded by the NSF as part of grant 1359275 at the University of Colorado Colorado Springs.

B. Methods

1) Data Collection: EEG data was collected from an 4x8 endodermal electrode array at a frequency of 1000 Hz (31 channels, one channel malfunctioned, Fig. 1). The data recording lasts

for 500 seconds, during which epileptic seizures were evoked using 4-aminopyridine and EEG data was recorded from the electrode array. The spiking voltage of each recorded electrode is used as a signal and is referred to as a channel (Fig. 2). Each channel has a continuous data series that spans the 500 seconds.

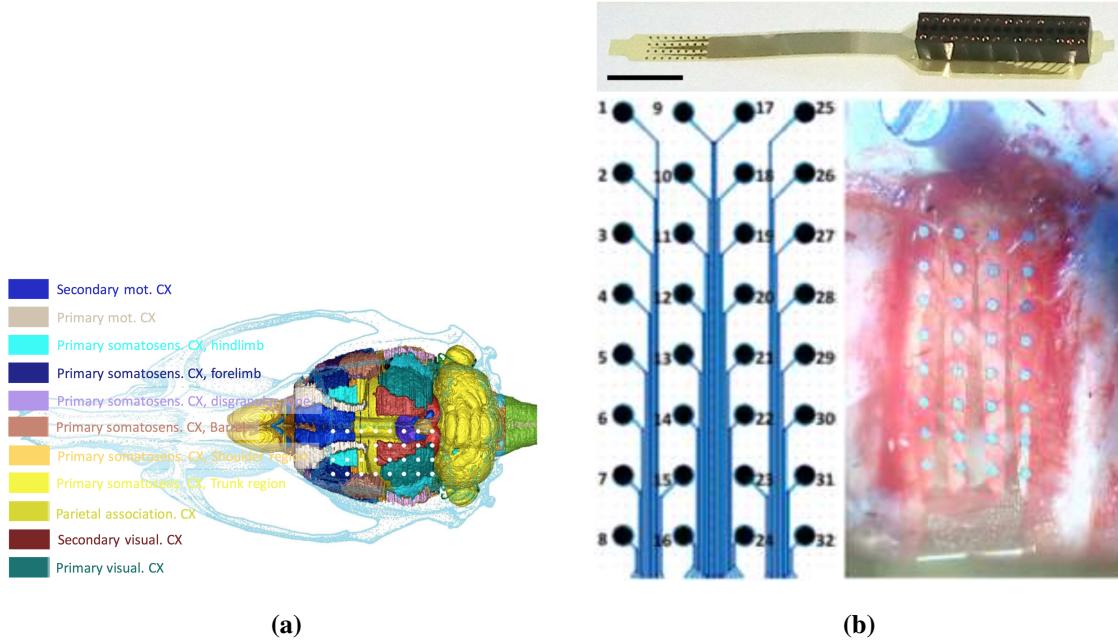


Fig. 1: (a) A composite drawing showing the brain in the scull. The 3D reconstruction of the brain has been made using the maps of the Paxinos atlas, and the localization of cortical areas are indicated by different colors. White points indicate the position of the recording sites of the membrane electrode. Names for the cortical areas are also shown (based on Hjornevik *et al.* and Paxinos *et al.* [1] [2]). (b) Photograph of a membrane electrode shows the construction on the top, the numbering (bottom right) and the surgical implantation (bottom left) is also shown. Electrode 1 malfunctioned during recording.

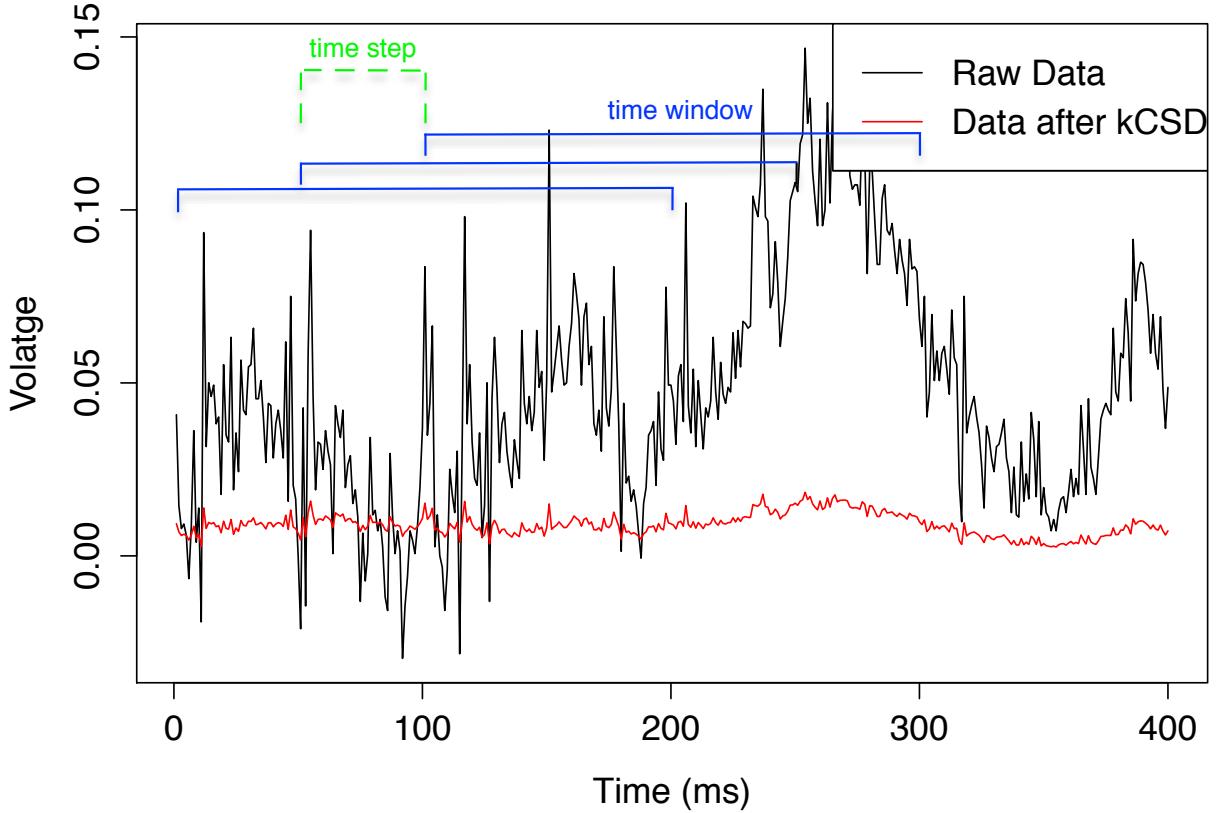


Fig. 2: Sample data from electrode 1, in both its raw form and then in its form after kCSD transformation. Applying kernel Current Source Density extracts the true relationship of neural activity, and in theory eliminates disturbances that might be caused by signal interference. Three sample blue lines are drawn to demonstrate the sliding windows that were used to calculate Sugihara Causality between brain regions. In each time window, the causality was calculated between each brain region, and the time window was moved by an amount *time step*. Different configurations were used for the time window and time step, as reported in table I. Here, the configuration shown has time window as 200 and time step as 50.

TABLE I: Configurations used for the sliding window when calculating Sugihara causality that was used as a measure of brain communication. These configurations inform our model of the granularity of brain communication. Since we do not know exactly how quickly information is being passed down from region to region, arbitrary choices were made to test how well each performs and make a more informed decision in future research.

Dataset name	Time window	Time step
dataset 50	200	50
dataset 200	200	200
dataset 250	2000	250

2) Data Preprocessing: Kernel Current Source Density (kCSD) method was used on the grid of channels to account for possible electrical interference with the direct measurement (Fig. 2) [3]. The measured potentials produced by kCSD arise as the linear combination of the transmembrane currents, which is a more direct and localized quantity to measure the neural activity. Therefore current source density distribution was calculated by the kCSD method and used for the analysis instead of the original EEG signals.

Afterwards, in order to reduce the data size that is operated on, we lumped channels from the same brain regions together by averaging them (Fig. 3). In addition to reducing dimensionality, this process also puts emphasis of causality on functional brain regions instead of a local cluster of neurons.

3) Sugihara Causality Measurement: After lumping the kCSD signals from the same region, we perform Sugihara Causality calculation on the regions by using a sliding windows method (Fig. 2). Because we do not know how granular of a time frame the brain uses to communicate (and how that affects the EEG signal we record), we use three different time scales as reported in table I. This produces 3 datasets with similar characteristics. The important thing to keep in mind is that after this step, each causality measure that resulted from the Sugihara Causality analysis is a measure of how much one region is affecting the other, and we use that as a way to describe brain communication and information flow. Each causality measure is therefore an *edge* of communication from one region to another, meaning that each brain region is a node and each causality measure is a weighted edge. Figure 4 shows a sample graph constructed from joining all the causality measures from a specific time window. This is essentially a (theoretical) snapshot of how each brain region was affecting other brain regions in that time window.



Fig. 3: The distribution and lumping of the brain regions in the brain. A total of 12 region channels were constructed from the initial 31 local channels. The schematic is based on rat brain atlas mapping.

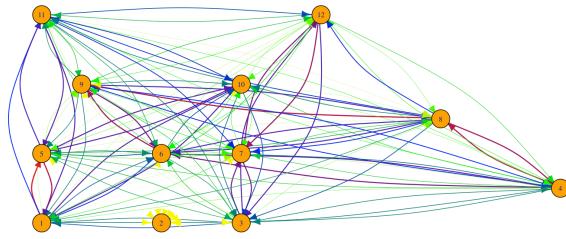


Fig. 4: Sample graph constructed from calculating the Sugihara causality between brain regions. Lumped brain regions correspond to the gray circular mapping in Fig. 3. Edge colors represent the strength of the causal relationship. From weakest to strongest: Yellow, Green, Blue, Red. The edge weight ranges from 0 to 1.

C. Problem Statement

Given a mesh of EEG signals that have been translated into a brain region communication graph series, can we cluster the graphs into meaningful segments that can help us find epilepsy states. To find out, we plan to use unsupervised clustering in the form of Gaussian Mixture algorithm. We also employ principal component analysis (PCA) to help reduce the dimensionality of the data for faster processing time.

D. Metrics

The unsupervised learning used here along with the high dimensionality of the dataset used in this project bring with them a unique difficulty in terms of defining a clear scoring method. In the scikit-learn library, there are only two methods that can be used to give a score to clusters (assuming no knowledge of *true* labels of each data point): Silhouette Coefficient and Calinsky-Harabaz Index. Both scores ascertain the quality of clustering by measuring how dense, well defined, and separated the clusters are. We are forced to use these metrics due to the nature of the data we are given, as we don't know the true labels of the data points. However, we do know the specific times of when epilepsy was induced, and since epilepsy is *one* state in the brain, then we can use the time indecies for epilepsy to be the one cluster we perform an accuracy (possibly F1) metric on.

II. ANALYSIS

A. Data Exploration and Visualization

The data we produced is a series of complete, directed graphs (all the same topology; each node is a brain region as seen in figure 1; each edge is a causality/communication measure) and their edge weights that we refer to as ρ scores. For each graph, there are 132 ρ scores, all of which range between 0 and 1 (decimals). Most of the ρ scores appear to be of high values, suggesting high connectivity between brain regions and therefore a high density graph (Fig. 5b). Furthermore, we also notice that many of the ρ scores to be similar within a pair of nodes (Fig. 6). For example, when looking at regions 1 and 2 within the first 200 milliseconds, they appear to be equally communicating to one another (Fig. 5a). This could be indicative of either bidirectional causality or unidirectional forcing, both of which can be detected by the Sugihara model.

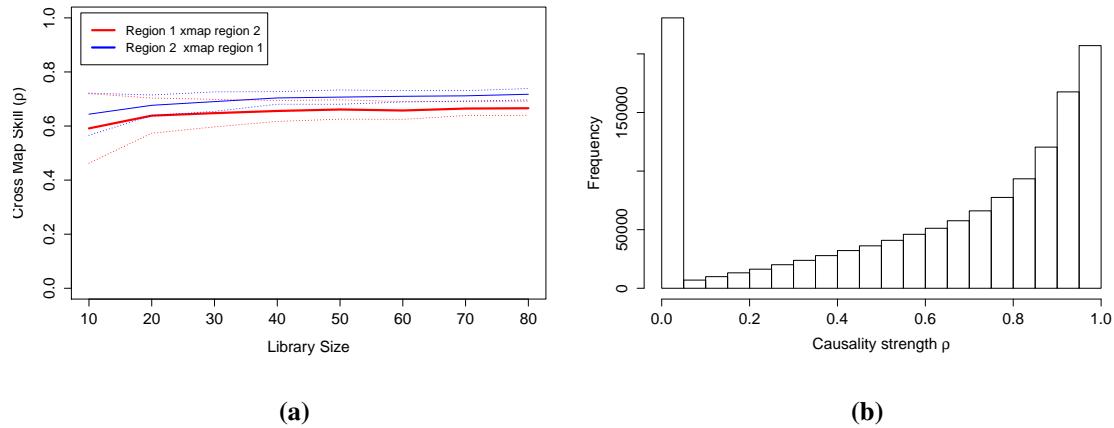


Fig. 5: (a) A sample graph plotting the convergent cross mapping skill (ρ) between region (nodes) 1 and 2 during the first 200 ms of the experiment. The skill mapping channel 1 to channel 2 is very similar to the one mapping channel 2 to channel 1. This might infer either bidirectional causality or unidirectional forcing. A similar pattern (close ρ value between pairs) was found for most of the pairs of edges. (b) The distribution of CCM skill (ρ) during the entire experiment. Many relationships appear to be causal in the brain, with equally as many being non-causal throughout the experiment. Causality was calculated from signals of lumped regions after calculating the CSD. Cross mapping was done on every pair of regions with library size of 80, and each pair has two causality directions. Sliding windows of 200 ms were used, with a sliding step of 50 ms.

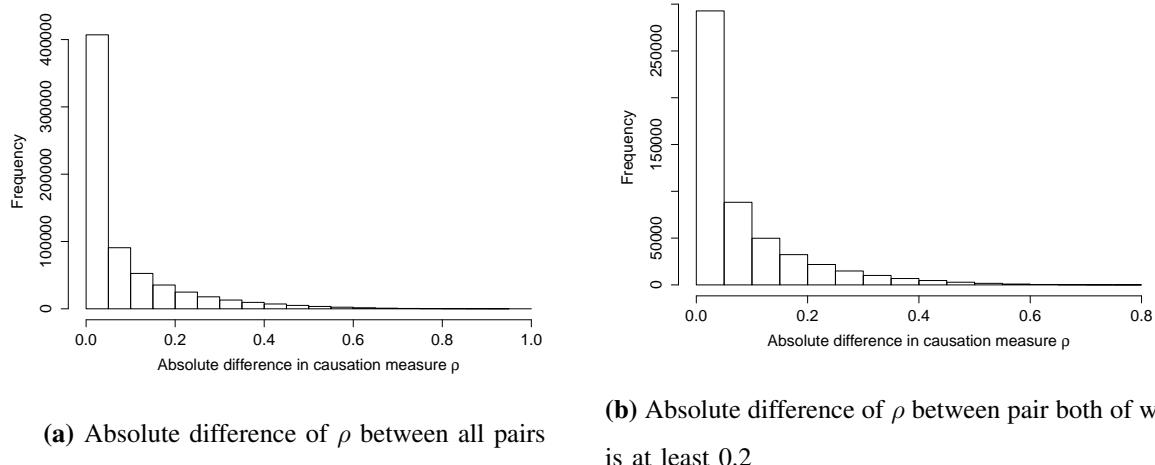


Fig. 6: Difference of causality values within a pair is small, even when accounting for non-existing relationships where both causalities are below 0.2. This similarity between directions of causality in a pair could imply a bidirectional relationship between most regions, or could alternatively imply a unidirectional forcing (synchrony) phenomenon. Sliding windows of 200 ms were used, with a sliding step of 50 ms. A library size of 80 was used.

B. Algorithms and Techniques

For the clustering, we choose the Gaussian Mixture algorithm as implemented by the Scikit learn python library. Gaussian Mixture Model (GMM) clustering can be thought of as a K-means algorithm that is generalized to account for the covariance of the data and centers of the Gaussian distributions. Because of this feature, while K-means algorithm only clusters data points in spherical shapes, the GMM model allows clusters to take many shapes. For these reasons, the GMM would be a more suitable algorithm to use in our case. GMM has a few parameters that could be tuned, the most important of which is the covariance type (full, tied, diagonal, or spherical) which dictates how the cluster edges are expanded and how the clusters grow. This parameter will be fine tuned using grid search.

C. Benchmark

Since this method has never been attempted before (using Sugihara Causality graphs to identify epilepsy), providing a succinct and complete benchmark is a difficult task. Furthermore, the clustering here is used to explore whether or not there are any extractable information from causal networks in the brain, and whether the brain communicates differently when it is suffering from epilepsy compared to when it is operating normally.

Keeping this in mind, we rely on the simple F1 score to provide a simple reliability measure of how well the clustering is separating and predicting the epilepsy cluster. Since we know that epilepsy doesn't happen often in our dataset (only 20 seconds of epilepsy related events in our 500 seconds dataset), we can assign the cluster with the least data points as the epilepsy one, and then measure how exclusive the epilepsy cluster is to the times when epilepsy was *actually* happening by measuring an F1 score. Since our epilepsy happens for about 4% of the dataset, then the baseline score we expect should be 96% of accuracy and 0.99 F1 score. These scores are attained if all the data points were labeled as one (non-epilepsy) cluster.

III. METHODOLOGY

A. Data Preprocessing

Some of our edge weights had some Not A Numerical values, and those were altered to 0, which is the lowest ρ score possible. Once that error was corrected and the data has been put in the graph series format, separated into the three different datasets, we perform Principal

Component Analysis (PCA) to reduce the dimensionality of the datasets. This both gives us insight into what components of the graph are logically important for the model and helps decrease the time required to run our model. The results of PCA was used to create three different datasets from each dataset: a dataset that covered 80% of the variance, a dataset that covered 90% of the variance, and a dataset that covered 99% of the variance. This was done to test how dimensionality reduction would affect the final clustering efforts. If the scoring for the reduced datasets with only 80% of the variance is similar to the scoring of the reduced datasets with 99% of the variance, then we use the less dimensional reduced datasets even though they give us a less accurate version of the original information. A list of the PCA datasets and their different variance levels are reported in table II. After the principal component analysis was done on the datasets, the components required to reach the threshold levels of 80, 90, and 99 percents were recorded in table III (Fig. 7).

TABLE II: Dataset names for the result of PCA analysis. Three different reduced datasets are produced each of which keeps a certain percentage of the original variance. Since we started with three datasets and created three new ones out of each by applying PCA with different levels of components, we arrive at 9 total reduced datasets.

	80% variance	90% variance	99% variance
dataset 50	reduced 50 80	reduced 50 90	reduced 50 99
dataset 200	reduced 200 80	reduced 200 90	reduced 200 99
dataset 250	reduced 250 80	reduced 250 90	reduced 250 99

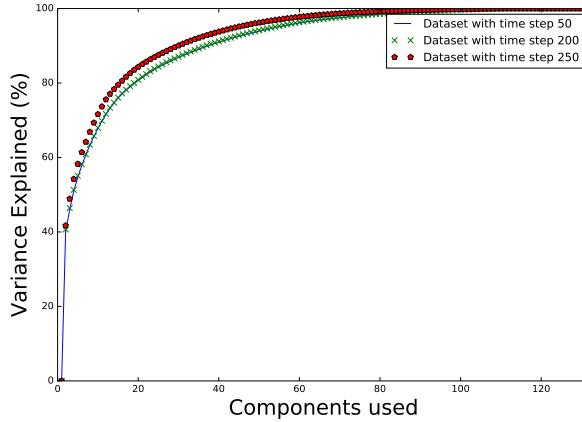


Fig. 7: Principle Component Analysis conducted on the three datasets reveals that the overall variance of the data can be maintained by using only a subset of the edges of the graph rather than all of them. The dataset with a time step of 250 ms needs less components to explain the variance than the other two, while the datasets with 200 and 50 ms time steps showed an almost exact level of variance explained with the number of components. This suggests that the 250 ms dataset is not complex enough, which might indicate that 2000 ms is too big of a time frame to record the granular communication speed of the brain.

TABLE III: Different component numbers are needed for each dataset to maintain a certain level of variance explained using PCA for dimensionality reduction. These values for components required for reaching a score standard were used to produce reduced forms of the datasets as noted in table II. These reduced forms of the datasets were then analyzed in figure 8.

	80% variance explained	90% variance explained	99% variance explained
Components in dataset 50	20	38	87
Components in dataset 200	20	37	87
Components in dataset 250	16	30	76

B. Implementation

Once the dimensionality of the data was reduced with PCA, a Gaussian Mixture clusterer was initiated for each dataset from table II, and all number of clusters from 2 to 40 were tried and tested by the Silhouette Coefficient and Calinsky-Harabaz Index scorers (Figs. 8, 9).

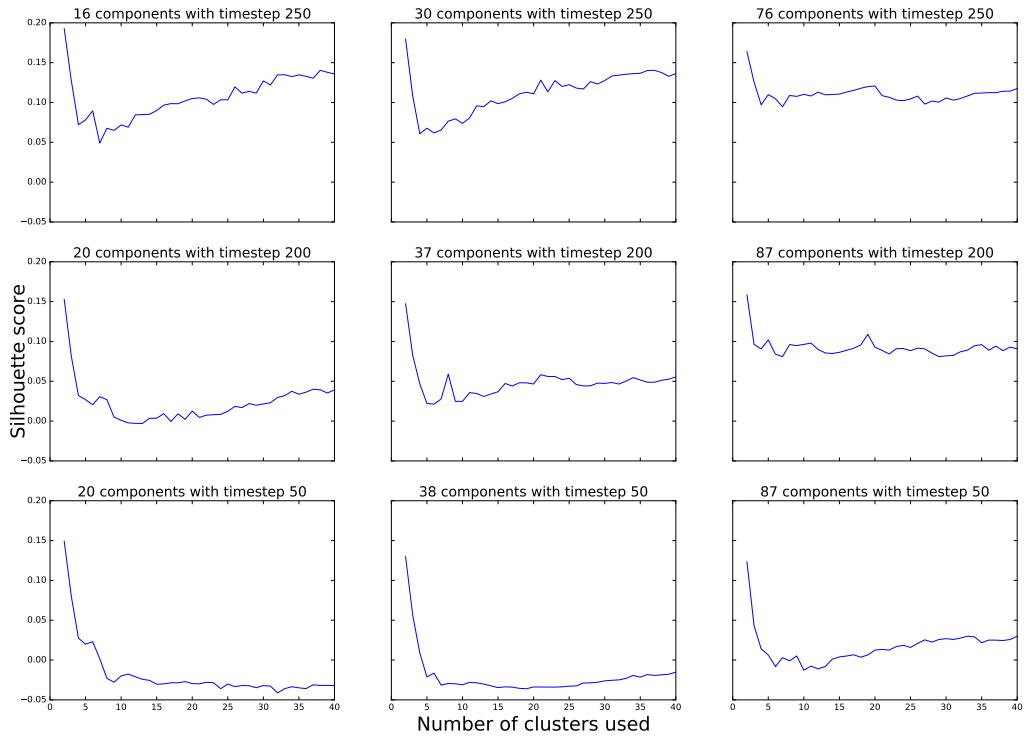


Fig. 8: Silhouette scores for clustering with different cluster numbers. Scores close to 1 suggest clusters that are dense and well separated; scores close to -1 suggest clusters that overlap and are not dense. The best score is consistently when there are only two clusters, and dips greatly when creating more clusters but begins to rise after 30 clusters. With the dataset 250 and dataset 50, PCA with less components (16 and 20 components respectively) showed a better silhouette score than the other, higher component reduced datasets. However, dataset 200 showed a better silhouette score when more components were used. Especially with a rather thin margin of difference in scores between components, it is difficult to conclude whether having more PCA reductionist approach is beneficial.

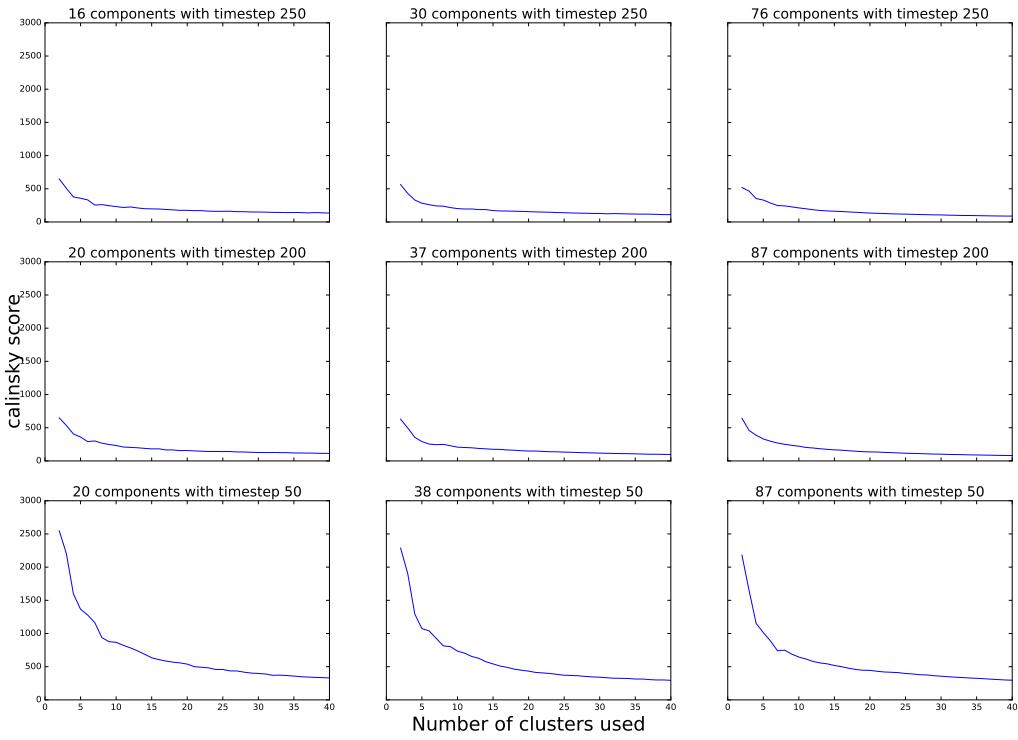


Fig. 9: Calinsky-Harabaz scores for clustering with different cluster numbers. Higher scores mean more separated and dense clusters, which is positive. The score dips greatly when creating more than two clusters, but begins to rise after 30 clusters. The silhouette scores show that having more components when reducing with PCA results in better clustering scores. The scores also show that scores for the reduced data that keep 99% of the variance (right column) are all similar in score and trend.

C. Refinement

From figures 8 and 9, it can be seen that the best clustering number is two, and it happens with dataset *reduced 50 80* (dataset that is made with time step of 50 and transformed with PCA). We therefore fine tuned a Gaussian Mixture model with the dataset 50, and arrived at a *covariance type* of 'tied' and a silhouette score of 0.254. This is an improvement by 0.05 in the silhouette score, which is a 25% improvement.

IV. MODEL EVALUATION AND VALIDATION

The results so far indicate that the rat's brain showed the presence of two distinct brain states, as we can cluster the brain's communication network into two clusters. This could be good

news, since we can assume that the two brain states are *epilepsy state* and *normal state*. This simplifies our choice of metric to evaluate the model since we can use a simple F1 measurement by assuming the less frequent brain state is the epileptic state, and record the less occurring cluster as the epileptic cluster; we then test the time indecies of this cluster with the time indecies we were given for when epilepsy was evoked. Since we know that the epilepsy happened between seconds 447-457 and 250-260, we set true labels to be 0 for all data points that were in that time period, and 1 for all other data points. We then set the less frequent cluster to be of label 0 (epilepsy cluster) and the more frequent cluster to 1 (non-epilepsy cluster). Our evaluation metric then becomes the F1 score of the predicted labels against the true labels that we produced given the knowledge of our experiment.

Using the fine-tuned model, we achieve an F1 score of 0.788, which seems acceptable at first, but is rather disappointing when realizing that a blind guess of setting all the labels to 1 (non epileptic state) results in an F1 score of 0.99. This is mainly due to the fact that the epilepsy state is infrequent, and so a clustering of non-epilepsy state as epilepsy creates a great error. This indicates that our model does not perform better than blindly guessing that there is only one state in the brain. This does not align with our hopes and expectations for the clustering solution.

V. CONCLUSION

A. Free-Form Visualization

To analyze what our model is doing, we plot an annotated time strip that shows what each time window of causality was clustered to (Fig. 10). From this graph, it can be seen that there is no correlation between what we observe in the real-life experiment and our clustering on the graph series constructed from Sugihara Causality measures.

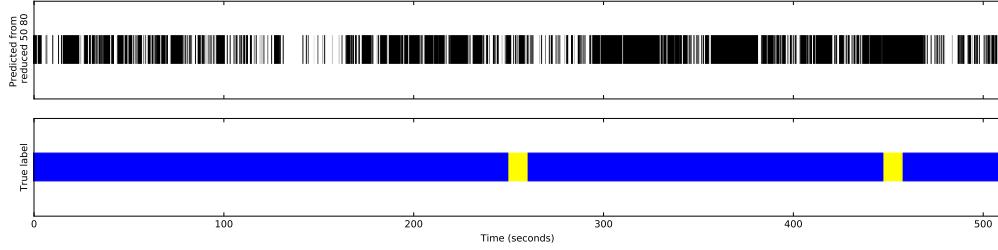


Fig. 10: Time strips that show the epilepsy time track and the results of the respective clustering done on the Sugihara Causality graphs from those time windows. Blue corresponds to true non-epileptic state; yellow to true epileptic state; white to predicted epileptic state; and black to predicted non-epileptic state. There are only two clusters in the top bar. The clustering results does not show any correlation with the actual epilepsy induction time series.

B. Reflection

This project transformed 32 EEG electrode signals into 12 regional neuronal activity signals, and created a (theoretical) communication network from those signals using Sugihara Causality. These communication graphs were analyzed by unsupervised learning using the Gaussian Mixture model for the goal of gaining some insight into the epileptic state, and test whether unsupervised clustering is able to discriminate between brain states from the communication networks. From our initial results, it is clear that our approach is not successful in distinguishing brain states in a rat's brain. However, the ground work has been laid for future explorations in the topic of communication networks in the brain.

C. Improvement

Since various improvements can be made to this work, we break it down by stage of implementation.

1) *Methods:* It would have been enriching to our knowledge of our true labels if we had access to exactly what the rat was doing during the experiment. A video capture of the rat, or recorded structured interventions to the rat's behavior could have provided more opportunity to label the data with ground truth. For example, we could have shone light into the rat's eyes, and recorded that as a separate brain state; we could have put the rat on a hamster wheel and recorded his running as a separate brain state; eating food could be a brain state; and sleeping is another possible brain state. All these interventions to the rat's environment could enable us to

gain more labels rather than the simple epilepsy, non-epilepsy labeling system we worked with. Although this complicates our approach, it would allow us access to better scoring metrics other than the silhouette and Calinksy-Harabaz scores, where the labels have to be known.

2) Data Processing: Apart from collecting the data and applying Sugihara Causality on the dataset (which took 2 days), calculating the current source density using kCSD was by far the most time consuming process. Even though we do not discuss the details of how that was done in this project (we provide already processed kCSD data), it would have been nice to see a GPU implementation of the kCSD, which took advantage of CPU parallelism but not GPU parallelism. This would massively improve the speed of the implementation and allow for testing more various approaches in the Sugihara Causality step.

3) Cluster Scoring with Higher Dimensions: Due to the high dimensionality of the dataset we worked with, it was difficult to visually measure how well the clustering with different number of clusters took place. Beyond three dimensions, it is impossible to visualize the dataset with labels, and although the PCA reduction resulted in a greatly reduced dimensional feature, it was not enough to allow us to visualize the clusters visually.

ACKNOWLEDGMENT

Special thanks goes to Dr. Somogivari and Dorottya R. from Wigner RCP, Budapest for providing the neuronal data used in this paper. Furthermore, Dr. Peter Erdi is thanked for his patient guidance of this research topic which was instrumental for the production of this paper.

REFERENCES

- [1] T. Hjornevik, T. B. Leergaard, D. Darine, O. Moldestad, A. M. Dale, F. Willoch, and J. G. Bjaalie, “Three-dimensional atlas system for mouse and rat brain imaging data,” *Frontiers in neuroinformatics*, vol. 1, p. 4, 2007.
- [2] G. Paxinos, C. Watson, P. Carrive, M. Kirkcaldie, and K. Ashwell, “Chemoarchitectonic atlas of the rat brain,” 2009.
- [3] J. Potworowski, W. Jakuczun, S. Łski, and D. Wójcik, “Kernel current source density method,” *Neural computation*, vol. 24, no. 2, pp. 541–575, 2012.