Categorizing Individual Neurons in the Brain using Low Pass Filtering and K-Means Clustering

by Isaac Donis

May 11, 2017

Abstract

Humans have about 100 billion neurons in their brain, with an individual neuron having up to 7,000 connections to other neurons. Knowing which neurons are connected within a given brain can improve the treatment of neurological disorders, as well as increasing general knowledge about the brain; but due to the sheer amount of cells, it is often difficult to map the connections within the entire human brain. One method of mapping the brain, called calcium imaging, attempts to model the connections between each neuron by making neurons light up when a calcium ion passes through each individual neuron cell in a path of cells. This method, which has been found to be accurate within the brains of zebrafish, lose accuracy with larger brains such as that of a human being. The proposed method in this paper does not necessarily map each individual neuron to another one, but rather assigns a neuron to a cluster using a clustering algorithm; each neuron is assigned to a specified "group," and neurons within the same group are classified as being within close proximity of each other, and thus two neurons in the same cluster have a higher probability of being connected to each other than two neurons in different clusters.

Background Information & Data

Biological neural networks are groups of neuron cells within the brain that together perform a certain function, such as walking or talking. While these actions are happening neurons are sending signals, or information, to one another. Visually seeing and recording the transfer of this information, therefore, is effectively mapping the neural network of the brain. One way of seeing information transfer between neurons is called fluorescence imaging, the process in which a genetically modified neuron lights up whenever a calcium ion goes through it. The brightness of these excited neurons can be recorded and quantified, thus giving a trail of where the ion goes as neurons illuminate and dim depending on if the ion is at the nucleus, or was recently there.

The data used in this experiment was simulated florescence data of calcium ions traveling within the brain of a zebrafish. The florescence level of a given neuron was taken once every 20ms for an hour, and a sample of 100 neurons is used in this study. The dataset is free and available to all on Kaggle.com [Kag14]. As a model for what the dataset looks like, the first 30 seconds of activity for one neuron is shown below in Figure 1, where the x-axis is time measured in seconds and the y-axis is the florescence level.

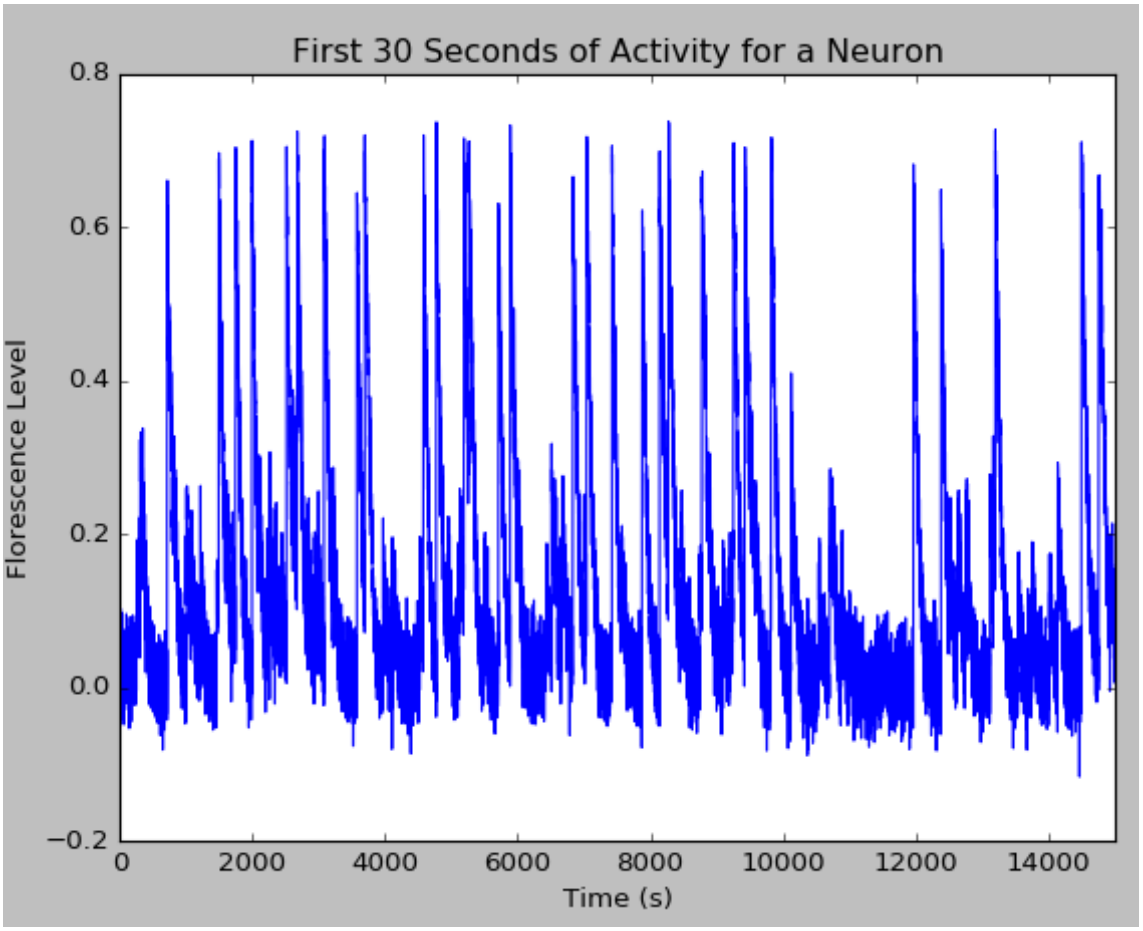

Figure 1: The first 30 seconds of activity for a sample neuron

The Approach

The approach in which this problem was solved was the following:

1. Load sample data for 100 neurons
2. Convolve the data with a differentiator, because taking the derivative of the data would pinpoint peaks and valleys in the data. This method was found in Signals & Systems (Oppenheim, Willsky, 1996) [OW96].
3. Pass the convolved data through a low pass filter, removing the noise in the signal
4. With a signal that is now easier to process, use K-Means clustering to organize which neurons are close in proximity. This process is outlined on Python's open-course website [Syp16].

As shown in Figure 1 above, the data for one neuron is extremely noisy. The important information is when the florescence level peaks, and all other data is the florescence level fluctuating and dimming as the ion moves farther away from the neuron. Intuitively, the farther away the calcium ion is from the neuron, the dimmer that neuron will be. Neurons within the same cluster should have similar brightness levels around the same time. In order to process the important information within the data, i.e, when the neuron is really bright or getting dimmer at a specific time, the data was smoothed with a differentiator in series with a low pass filter, as shown in Figure 2. This allowed for the peaks and valleys of each individual neuron signal to be identified and compared

easily. The order and cutoff frequency for the filter was found through trial and error by visually determining when the peaks and valleys would be most defined.

After each neuron sample was made smooth by the filter, the data was ready to be processed through a clustering algorithm. In the specific type of clustering used in this study, K-Means clustering, each neuron is assigned to a "cluster" or group. In this study, there were 19 groups, mimicking the 19 parts of the zebrafish's brain, as outlined in Neuroanatomy of the Zerbafish Brain: A Topological Atlas, [MFW12], and shown in Figure 3.

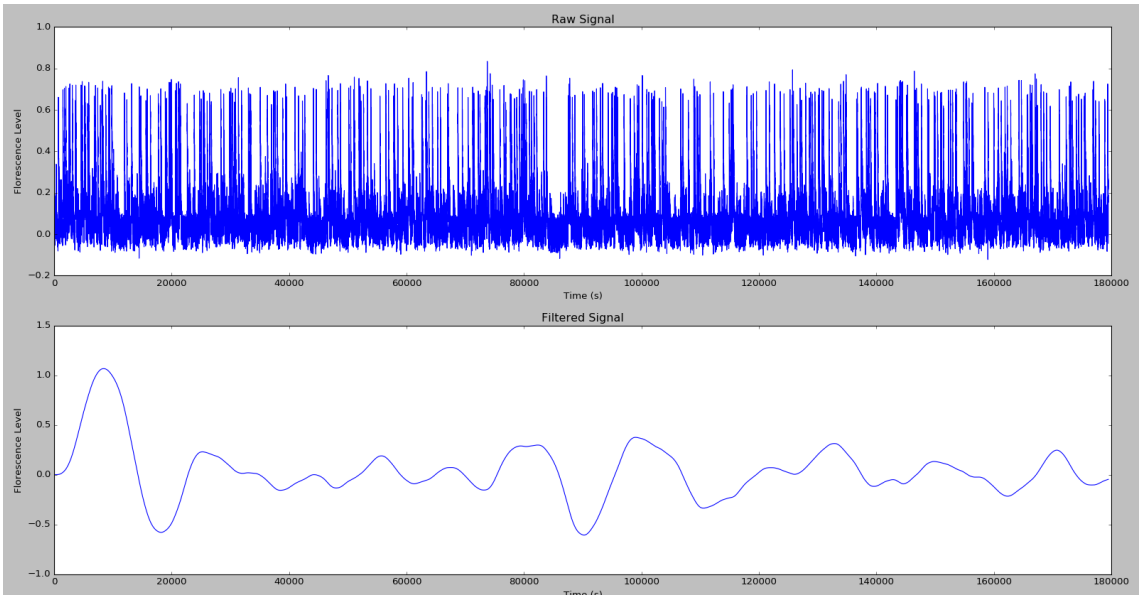

Figure 2: Example of a raw neuron data signal and a filtered one

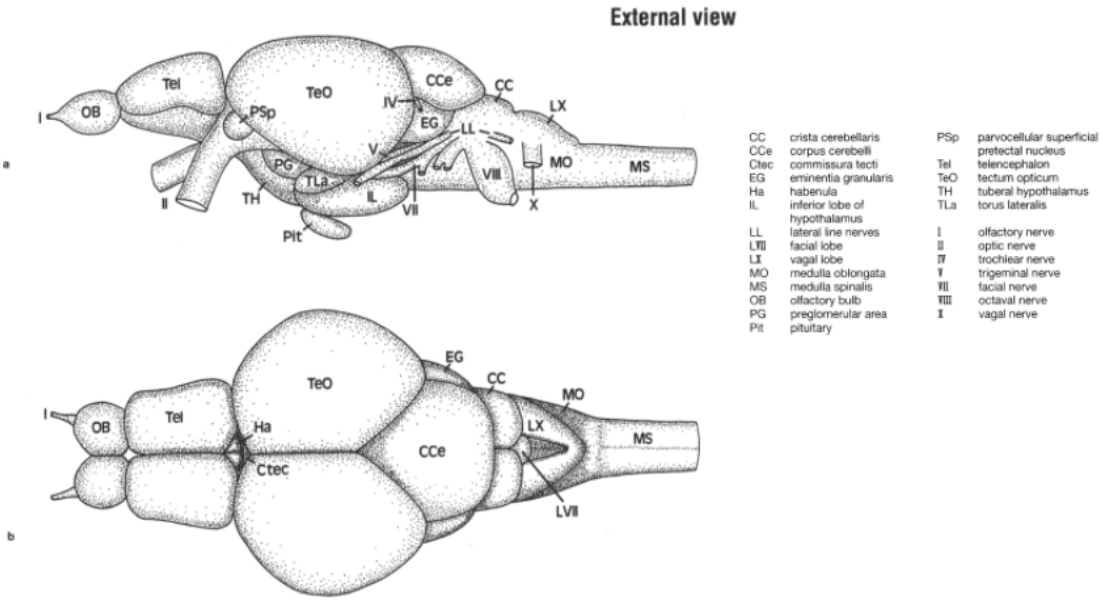

Figure 3: Parts of a zebrafish brain

Results

Each neuron was assigned to a cluster ranging from 0 to 18, and the distribution of this assignment is shown in Figure 4, shown below:

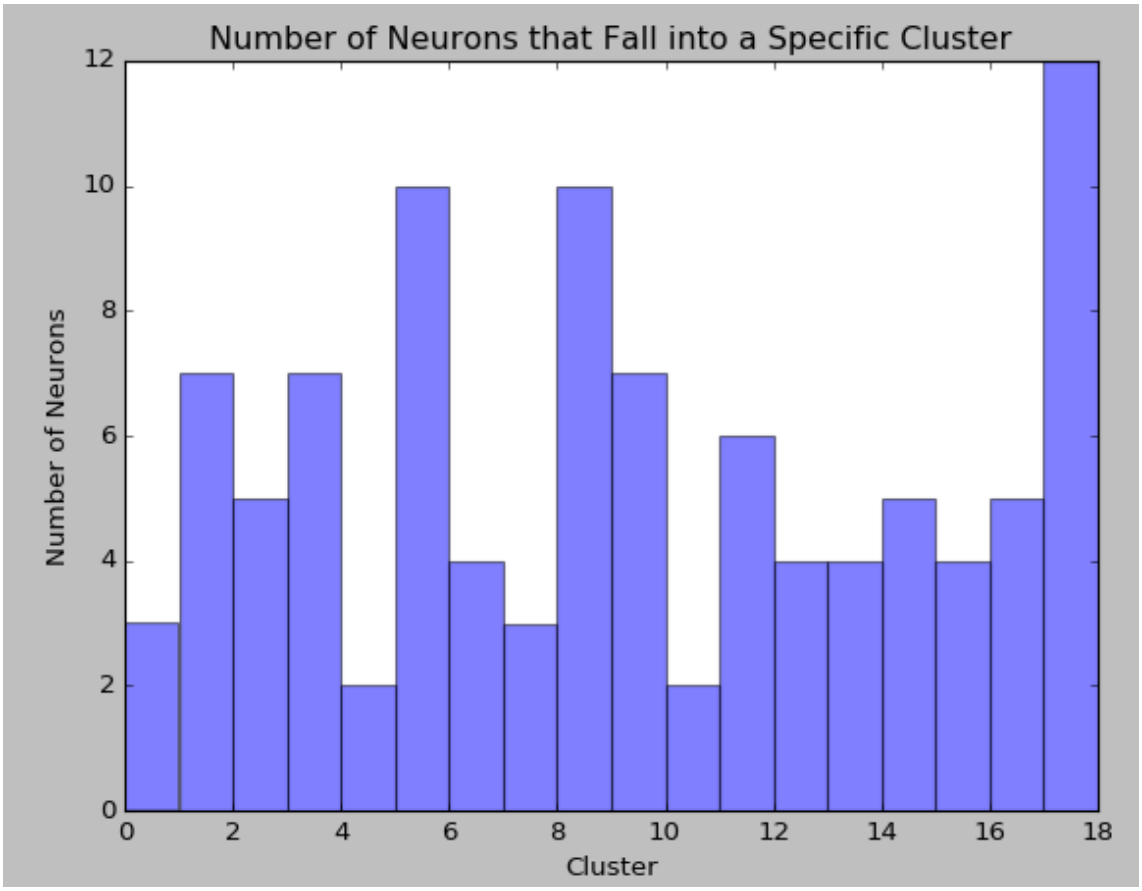

Figure 4: Distribution of the neurons being assigned to clusters

The full results of which neurons were correlated to which cluster is shown in Table 1:

Neuron 1	8		Neuron 26	10		Neuron 51	1		Neuron 76	13
Neuron 2	8		Neuron 27	8		Neuron 52	9		Neuron 77	12
Neuron 3	17		Neuron 28	13		Neuron 53	15		Neuron 78	1
Neuron 4	8		Neuron 29	7		Neuron 54	2		Neuron 79	9
Neuron 5	17		Neuron 30	14		Neuron 55	6		Neuron 80	2
Neuron 6	5		Neuron 31	8		Neuron 56	11		Neuron 81	3
Neuron 7	0		Neuron 32	9		Neuron 57	2		Neuron 82	3
Neuron 8	13		Neuron 33	0		Neuron 58	12		Neuron 83	16
Neuron 9	1		Neuron 34	17		Neuron 59	2		Neuron 84	11
Neuron 10	17		Neuron 35	5		Neuron 60	5		Neuron 85	14
Neuron 11	17		Neuron 36	14		Neuron 61	6		Neuron 86	18
Neuron 12	7		Neuron 37	0		Neuron 62	1		Neuron 87	8
Neuron 13	8		Neuron 38	15		Neuron 63	17		Neuron 88	16
Neuron 14	5		Neuron 39	8		Neuron 64	14		Neuron 89	14
Neuron 15	11		Neuron 40	3		Neuron 65	5		Neuron 90	3
Neuron 16	7		Neuron 41	3		Neuron 66	5		Neuron 91	4
Neuron 17	3		Neuron 42	3		Neuron 67	11		Neuron 92	17
Neuron 18	2		Neuron 43	12		Neuron 68	13		Neuron 93	9
Neuron 19	8		Neuron 44	5		Neuron 69	5		Neuron 94	12
Neuron 20	9		Neuron 45	5		Neuron 70	17		Neuron 95	1
Neuron 21	11		Neuron 46	6		Neuron 71	16		Neuron 96	17
Neuron 22	5		Neuron 47	15		Neuron 72	10		Neuron 97	16
Neuron 23	9		Neuron 48	1		Neuron 73	17		Neuron 98	17
Neuron 24	1		Neuron 49	15		Neuron 74	11		Neuron 99	4
Neuron 25	8		Neuron 50	9		Neuron 75	6		Neuron 100	16

Figure 5: Neurons and their clusters

Conclusion & Further Work

In conclusion, the algorithm was able to successfully find which neuron was correlated to which cluster. Each cluster represented a different part of the zebrafish's brain, and so a reasonable extension of this study would be to map each cluster number to a specific part of the brain. With more computational power, a larger sample size than 100 neurons can also be studied.

With regards to the filter was cleaned the data before it was processed for clustering, there are other algorithms that calculate the optimal filter order and cutoff frequency. These methods were not employed here, but would optimize the filter used.

References

[Kag14] Kaggle. Connectomics, 2014.

[MFW12] Heinrich Reichert Mario F. Wulliman, Barbara Rupp. Neuroanatomy of the zebrafish brain: A topological atlas. page 19, 2012.

[OW96] Alan V. Oppenheim and Alan S. Willsky. *Signals & Systems*. Prentice Hall, 1996.

[Syp16] Sypi.org. K-means clustering and vector quantization, 2016.