Crabsort

Spike-sorting for small circuit networks

Alec Hoyland Cosmo Guerini

Center for Systems Neuroscience

June 25, 2019

Spike-sorting primer

Spike-sorting is the process of mapping action potentials to the originating cell.

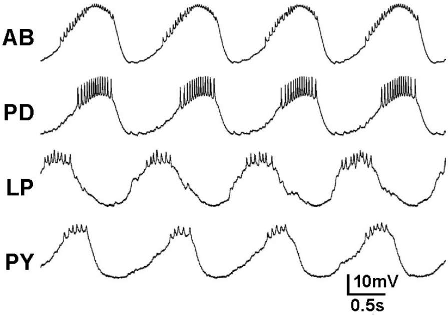

Figure: Intracellular recordings of pyloric cells.

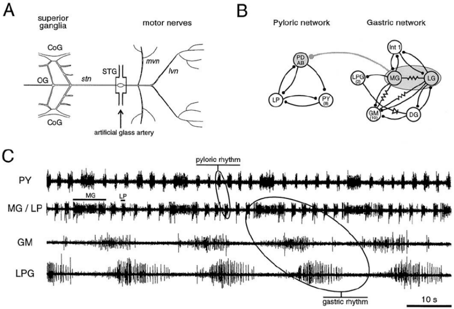

Figure: **A** circuit diagram of nerves; **B** connectivity diagram, circles are cells, synapses are lines and dots; **C** extracellular recording of motor nerves.

How to sort spikes

1. Identify spikes from membrane potential waveform (easy)
2. Sort spikes using some magic algorithm (hard)

For large networks, spike sorting means:

1. Methods: PCA, SVD, stochastic k-means matching of correllograms
2. Data: 100s of channels but without a ground truth

For small networks, spike sorting means:

1. Methods: dimensionality reduction, machine learning
2. Data: few channels with known activity

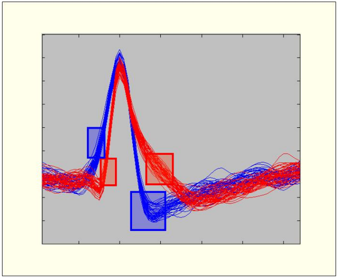

Figure: Spike-sorting using PCA and window filtering by spike-waveform analysis (R. Quian Quiroga).

Crabsort

Dimensionality Reduction

Dimensionality reduction is the process of taking high-dimensional data and representing it in a lower dimensional space.

Crabsort allows the user to:

1. Dimensionally reduce the data to a 2-dimensional manifold
2. Interactively label the data

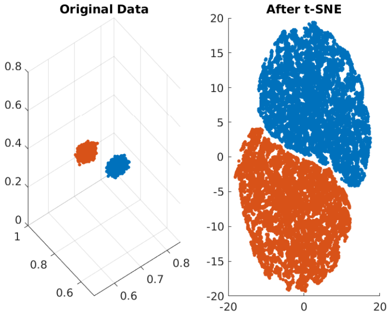

Figure: Using the t-SNE algorithm to reduce a 3-dimensional dataset to a 2-dimensional dataset.

t-distributed stochastic neighborhood embedding

1. Choose nearest neighbors for x_i via a multi-dimensional Gaussian probability distribution centered at x_i.
2. Compute t-distributed similarity measure in low-dimensional space.
3. Minimize the Kullback-Leibler divergence between the high-dimensional distribution and the low-dimensional distribution.

With FFT-interpolation

1. Computing objective function requires a convolution. So, interpolated equispaced grid, and compute convolution in Fourier space using FFT.
2. Use approximate nearest neighbors (ANNOY) to multithread kNN task.
3. 15-30x faster at $\mathcal{O}(n \log n)$.

Uniform manifold approximation and projection

1. k-nearest neighbors leads to a weighted graph which defines a manifold
2. Spectral clustering on the Laplacian matrix
3. Minimization of fuzzy simplicial set cross entropy optimizes embedding.

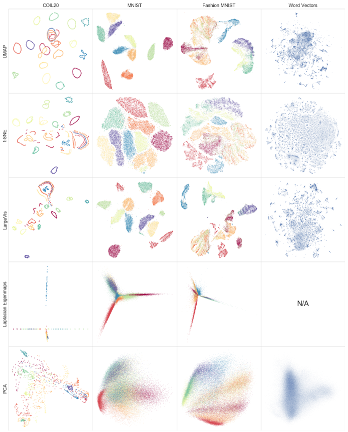

Figure: Comparison of UMAP, t-SNE and other dimensional reduction techniques.