

ToMATo clustering algorithm for spike sorting

Louise Martineau | University of Strasbourg

Recording and decoding the activity of multiple neurons is a major subject in contemporary neuroscience. Extracellular recordings with multi-electrode arrays is one of the basic tools used to that end. The raw data produced by these recordings are almost systematically a mixture of activities from several neurons. In order to find the number of neurons which contributed to the recording and identify which neuron generated each of the visible spikes, a pre-processing step called spike sorting is required. Spike sorting is nowadays a semi-automatic process that involves many steps. Indeed, following some initial steps (data normalization, spike detection, events construction), spike sorting boils down to a clustering problem in high dimension. It is therefore accompanied most of the time by a dimension reduction, since classic clustering algorithms do not perform well in high dimension. This dimension reduction step is sensitive to the presence of event superposition (akin to outliers) that lead to poor clustering results. Neuroscientists are then usually led to perform an extra pre-processing step to remove these superpositions; this step is not completely automatized and does a much better job when supervised by an expert in spike sorting. The use of the ToMATo clustering algorithm helps to simplify and streamline this whole spike sorting procedure. \\



The ideas of a graph-based mode-seeking were already introduced in the algorithm of [Koontz et al., 1976, IEEE Trans Comput.,C-25:936] on which ToMATo is based. The ToMATo algorithm is thus usable even in high dimension. Addressing the first issue, methods such as the Mean-Shift try to smooth $\frac{1}{5}$, but require a smoothing parameter choice. The innovative approach of ToMATo resides in the use of persistence theory. With persistence, a notion of prominence of peaks can be defined, and prominent peaks of $\frac{1}{5}$ correspond to prominent peaks of $\frac{1}{5}$. Clusters found by mode-seeking are merged together so that the final clusters correspond only to prominent peaks of the true density $\frac{1}{5}$, and not some spurious, noise induced, peaks of $\frac{1}{5}$. The correct number of clusters is therefore recovered, explaining the effectiveness of the ToMATo method. $\$

We show that ToMATo can be easily applied for spike sorting, enabling to reduce the numbers of steps typically involved in this procedure. Indeed, since it can be applied directly to high dimensional data (in our application the data is \$180\$-dimensional), there is no need for dimension reduction before clustering, making the sensitivity to superpositions disappear. Moreover, ToMATo provides a very easy way of choosing the right number of clusters, solving thereby a significant problem in clustering. Finally, the algorithm is easy to use even for non specialists of Topological Data Analysis. Our use case demonstrate the spectacular performances of this approach. \\

Supported by the ANR: project ANR-22-CE45-0027.