Optimal Interval Clustering: Application to Bregman Clustering and Statistical Mixture Learning

IEEE Signal Process. Lett. 21(10): 1289-1292 (2014). BIBTeX:`J2014-OptimalIntervalClustering` [1]

Let $\mathbb{X}$ be a one-dimensional space totally ordered with respect to $<$, and $\mathcal{X} = \{x_1, ..., x_n\} \subset \mathbb{X}$ a set of n distinct (weighted) elements. Let us sort $\mathcal{X}$ in $O(n \log n)$ time, so that we assume $x_1 < ... < x_n$. An *interval clustering* of $\mathcal{X}$ into $k \in \mathbb{N}$ clusters partitions $\mathcal{X}$ into pairwise disjoint subsets $\mathcal{C}_1 \subset \mathcal{X}, ..., \mathcal{C}_k \subset \mathcal{X}$ so that $\mathcal{X} = \biguplus_{i=1}^k \mathcal{C}_i$: $\underbrace{[x_{l_1=1} ... x_{r_1=l_2-1}]}_{\mathcal{C}_1} \underbrace{[x_{l_2} ... x_{r_2=l_3-1}]}_{\mathcal{C}_2} ... \underbrace{[x_{l_k} ... x_{r_k=n}]}_{\mathcal{C}_k}$. The output is a collection of k intervals $I_i = [x_{l_i}, x_{r_i}]$ that can be encoded using $k-1$ *delimiters* l_i ($i \in \{2, ..., k\}$). To define an optimal clustering among the $\binom{n-1}{k-1}$ different contiguous partitions, we ask to minimize a clustering *objective function* $\min_{l_1=1<l_2<...<l_k} e_k(\mathcal{X}) = \bigoplus_{i=1}^k e_1(\mathcal{C}_i)$, where $\bigoplus$ is a commutative and associative operator. We present a $O(n^3 k)$-time generic dynamic programming method to compute the optimal 1D interval clustering that includes 1D Euclidean k-means, Bregman k-means, k-medoids, k-medians, k-centers, etc. The dynamic programming requires $O(nk)$ memory to backtrack the optimal solution. For Bregman k-means, we reduce the complexity to $O(n^2 k)$ time by preprocessing cumulative sums of the elements of $\mathcal{X}$, and show how to include cluster size constraints. As an application, we report a learning algorithm for singly-parametric statistical mixtures maximizing the complete likelihood (k-MLE) that also performs model selection. We present experimental results on isotropic Gaussian mixtures and give necessary conditions on the family of parametric distributions that yields interval clustering: Namely, we require the connected property of the further maximum likelihood Voronoï diagrams (satisfied by singly-parametric exponential family mixtures).

References

[1] Frank Nielsen and Richard Nock. Optimal interval clustering: Application to Bregman clustering and statistical mixture learning. *IEEE Signal Process. Lett.*, 21(10):1289–1292, 2014.