Topological Approach to Hierarchical Clustering

September 24, 2019

References

Carlsson and Memoli, Classifying Clustering Schemes, Foundations of Computational Mathematics, vol. 13, no. 2, pp. 221-252

See

Persistent Sets and Dendrograms

Let X be a finite metric space. Let $\mathcal{P}(X)$ be the set of partitions of X, where a *partition* is a disjoint covering of X by subsets.

A *persistent set* P is a pair (X, θ) where $\theta : \mathbf{R}_{\geq 0} \to \mathcal{P}(X)$ is a function with the property that:

- if $r \leq s$, then $\theta(r)$ is a refinement of $\theta(s)$
- for any r, there is an $\epsilon > 0$ so that $\theta(r + x) = \theta(r)$ for all $x \in [r, r + \epsilon]$.

If, for large enough r, $\theta(r)$ is the partition with a single block equal to X, we say P is a dendrogram.

Linkage functions

Let ℓ be a real-valued function on pairs of subsets of X. We think of ℓ as measuring the distance between two set; such a function is called a linkage function.

- single linkage: $\ell(B, B')$ is the distance between the closest points of B and B',
- complete linkage: $\ell(B, B')$ is the distance between the farthest points of B and B',
- average linkage: $\ell(B, B')$ is the average distance between pairs of points in B and B'
- Ward's criterion: $\ell(B, B')$ is the change in the variance resulting from merging B and B'.

Clustering from linkage

In traditional clustering, one finds the "closest clusters" and merges them. This is sensitive to choices. A more canonical approach is the following.

For each $r > 0$, define an equivalence relation $\sim_{\ell,r}$ by saying two blocks A and B are r-equivalent if there is a sequence of blocks starting at A and ending at B with the distance between successive blocks at most r.

Let Θ_1 be the partition of X into single points. Inductively define Θ_{i+1} by letting r_i be the minimum linkage distance between distinct blocks in Θ_i, and then letting Θ_{i+1} be the equivalence classes of Θ_i under the r_i equivalence relation.

Define a function θ that jumps from Θ_i to Θ_{i+1} at r_i and is constant between the r_i.

Functoriality

- Let $\mathcal{M}$ be the category of finite metric spaces with injective, distance non-increasing maps.
- Persistent sets form a category where the requirement is that if $f : X \to Y$ then $f^{-1}(\theta_Y(r))$ is a refinement of $\theta_X(r)$ for all r.

Carlsson-Memoli ask what it means to require that a clustering method is "functorial". In other words, how do the clusters behave under maps between the underlying spaces.

Main Result

Let H be any clustering method which is

- functorial from metric spaces to persistent sets;
- preserves the underlying set, in the sense that if β is the forgetful functor from persistent sets to the underlying set, and α is the forgetful functor from metric spaces to sets, then $\beta H = \alpha$.
- Carries the two point metric space, whose points are at distance δ, to the persistent set that is the trivial partition for $r < \delta$ and the one-block partition for $r \geq \delta$.
- For any X, let s be the minimum distance among any pair of points in X (this is called the separation of X). Then $H(X)$ is the trivial partition for all $t < s$.

Then H is single linkage clustering.

What goes wrong in other cases

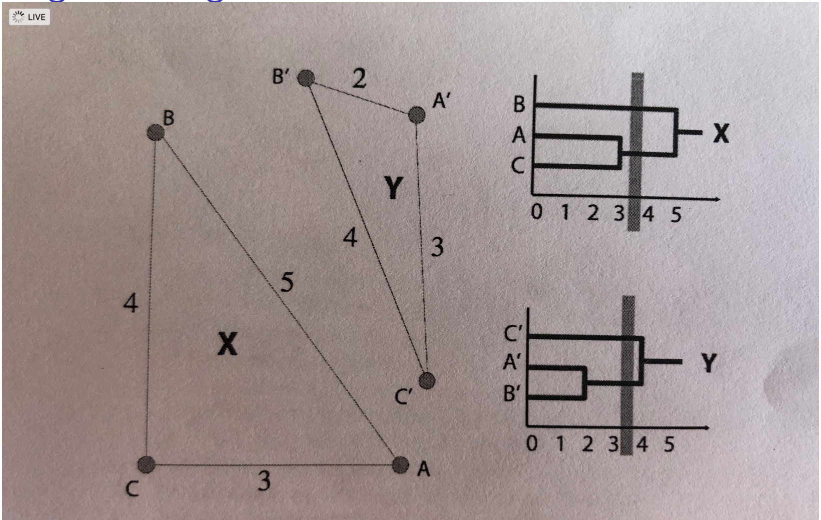

FIGURE 5. An example that shows why complete linkage fails to be functorial on $\underline{\mathcal{M}}^{inj}$ (and therefore also on $\underline{\mathcal{M}}^{gen}$).

Figure 1: From Carlsson-Memoli