Clusters as Functors

Tim Pollock

June 14, 2018

Why Functors

We wish to describe clustering algorithms as functors. There are multiple reasons for this:

- Intuitively, clustering algorithms should have some notion of consistency; how one dataset gets clustered should be related to how a similar dataset gets clustered.
- Category theory provides a useful framework to state properties and prove theorems, and fits naturally with other areas of math.
- Functorial schemes allows one to analyze qualitative geometric properties of the data, and has proven useful in many constructions, such as zig zag diagrams.

The Input Category $\mathbb{M}$

To define a clustering scheme $\mathfrak{C}$ as a functor, we need to specify an input category and an output category. It's clear that the objects of the input category should be finite metric spaces, but there are many choices of morphisms. We consider $\mathbb{M}^{iso} \subset \mathbb{M}^{inj} \subset \mathbb{M}^{gen}$, the categories with objects finite metric metric spaces and morphisms isometries, injective distance-non-increasing maps, and general distance-non-increasing maps respectively. The more morphisms, the stricter functorality is as a property.

The Output Category $\mathbb{O}$

As a clustering algorithm takes partions the points in a metric space into clusters, the output category $\mathbb{O}$ should have as objects pairs (X, P_X), where X is a finite set and P_X is a partion of X. To have functorality we will want maps of metric spaces to induce maps of clusters. Thus define a morphism in $f : (X, P_X) \longrightarrow (Y, P_Y)$ in $\mathcal{O}$ to be a map $f : X \longrightarrow Y$ such that P_X refines $f^{-1}(P_Y)$.

The Definition of a Clustering Functor

A *(standard) clustering functor* (scheme,algorithm) is a functor $\mathfrak{C} : \mathbb{M} \longrightarrow \mathbb{O}$ such that $\mathfrak{C} \circ \beta = \alpha$, where $\alpha : \mathbb{M} \longrightarrow$ Set and $\beta : \mathbb{O} \longrightarrow$ Set are the forgetful functors. Here $\mathbb{M}$ denotes any of $\mathbb{M}^{iso}, \mathbb{M}^{inj}$, or $\mathbb{M}^{gen}$.

The Vietoris-Rips Clustering Functor

For a fixed $\delta > 0$, we define the *Vietoris-Rips Clustering Functor* $\mathfrak{R}_\delta$ by setting $R_\delta(X, d_X) = X / \sim_\delta$ where $x \sim_\delta x'$ if there exists a path $x = x_0, \ldots x_k = x'$ such that $d_X(x_i, x_{i+1}) \leq \delta$.

Theorem

Assume $\mathfrak{C} : \mathbb{M}^{gen} \longrightarrow \mathbb{O}$ *is a clustering scheme such that there exists* $\delta_{\mathfrak{C}} > 0$ *with* $\mathfrak{C}(\Delta_2(\delta))$ *has one cluster iff* $\delta < \delta_{\mathfrak{C}}$. *Then* $\mathfrak{C} = \mathfrak{R}_{\delta_{\mathfrak{C}}}$.

Excisive Clustering Functors

We consider an intuitively desirable property for a clustering algorithm to have.

We will call a clustering functor $\mathfrak{C}$ *excisive* if $\mathfrak{C}(X, d_X) = (X, \{X_\alpha\})$ implies that $\mathfrak{C}(X_\alpha, d_X|_{X_\alpha \times X_\alpha}) = (X_\alpha, \{X_\alpha\})$. For instance, the Vietoris-Rips functor is clearly excisive.

Representable Functors

We wish to classify excisive functors. To do this we note that excision is equivalent to existence of a certain kind of explicit generative model. We say that clustering functor $\mathfrak{C}$ represented by $\Omega \subset \mathbb{M}$, written $\mathfrak{C}^{\Omega}$, if in $\mathfrak{C}^{\Omega}(X, d_X)$, points $x, x' \in X$ are in the same block if there exists:

- a sequence of points $x = x_0, \ldots x_k = x'$
- a sequence of metric spaces $\omega_1, \ldots \omega_k \in \Omega$
- a pair of points $(a_i, b_i) \in \omega_i$ and morphisms $f_i : \omega_i \longrightarrow X$ such that $f_i(a_i) = x_{i-1}$ and $f_i(b_i) = x_i$.

For instance, $\mathfrak{R}_\delta = \mathfrak{C}^{\{\Delta_2(\delta)\}}$

Change of Metric

Given a finite set X and a symmetric function $W_X : X \times X \longrightarrow \mathbb{R}_+$ we define a metric (called the *subdominant ultrametric*) by

$$\mathcal{U}(W_X)(x, x') = \min\{\max_i W_X(x_i, x_{i+1}) : x = x_0, x_1 \ldots x_k = x'\}$$

Given a set $\Omega \subset \mathbb{M}$ we define $\mathfrak{I}^\Omega : \mathbb{M} \longrightarrow \mathbb{M}$ which sends (X, d_X) to (X, d_X^Ω) where $d_X^\Omega) = \mathcal{U}(W_X)$ and

$$W_X(x, x') = \inf\{\gamma > 0 : \exists \omega \in \Omega, \phi \in \mathrm{Hom}(\gamma \cdot \omega, X), x, x' \in \phi(\omega)\}.$$

Classification of Finitely-Representable Functors

Theorem

Let $\mathbb{M}$ *be either* $\mathbb{M}^{gen}$ *or* $\mathbb{M}^{inj}$ *and* $\mathfrak{C} : \mathbb{M} \longrightarrow \mathbb{O}$ *finitely-represented by* $\Omega \in \mathbb{M}$*. Then* $\mathfrak{C} = \mathfrak{R}_1 \circ \mathfrak{I}^{\Omega}$*.*

Classification of Finitely-Representable Functors

Proof. We'll write $\mathfrak{C}(X, d_X) = (X, P)$ and $\mathfrak{R}_1 \circ \mathfrak{I}^\Omega(X, d_X) = (X, P')$. Let $x, x' \in B \in P$, $x = x_0, \dots x_k = x'$ be a sequence in X and $\omega_1, \dots \omega_k$ a sequence in Ω with $f_i \in \mathrm{Hom}(\omega_i, X)$ $f_i(a_i) = x_i, f_i(b_i) = x_{i+1}$ for some $a_i, b_i \in \omega_i$. Then $W_X^\Omega(x_i, x_{i+1}) \leq 1$ for all i, and thus

$$d_X^\Omega(x, x') \leq \max W_X^\Omega(x_i, x_{i+1}) \leq 1.$$

It follows that x and x' are in the same block of P'.

Classification of Finitely-Representable Functors

Let $x, x' \in B' \in P'$. Thus there exists a sequence $x = x_0, \dots x_k = x'$ such that $d^{\Omega}_X(x_i, x_{i+1}) \leq 1$. We deduce that for each i there exists a sequence $x_i = y^{(i)}_0, \dots y^{(i)}_{k_i} = x_{i+1}$ such that $W^{\Omega}_X(y^{(i)}_j, y^{(i)}_{j+1}) \leq 1$. Concatenating these lists we get a list $x = z_0, \dots z_N = x'$ such that $W^{\Omega}_X(z_l, z_{l+1}) \leq 1$. Thus there exists $\gamma_l \in (0, 1]$, $\omega_l \in \Omega$, and $f_l \in \mathrm{Hom}(\gamma_l \cdot \omega_l, X)$ with $z_l, z_{l+1} \in f_l(\gamma_l \cdot \omega_l)$. But then f_l is also a morphism $\omega_l \longrightarrow X$, and thus x and x' lie in the same block of P. $\square$

Scale Invariance

It's easy to see that the Vietoris-Rips clustering functor is not scale-invariant. Are there any useful clustering algorithms that are scale-invariant?

Scale Invariance

Theorem

If $\mathfrak{C} : \mathbb{M}^{gen} \longrightarrow \mathbb{O}$ *is a scale-invariant clustering scheme, then either:*

- $\mathfrak{C}$ *partitions any* (X, d_X) *into a single block, or*
- $\mathfrak{C}$ *partitions any* (X, d_X) *into singletons.*

Scale Invariance

Proof. Case 1: $\mathfrak{C}(\delta_2(1))$ is in one cluster. Then by scale invariance $\mathfrak{C}(\Delta_2(\delta))$ is in one cluster for all $\delta > 0$. Fix a $(X, d_X) \in \mathbb{M}^{gen}$ and let $\mathfrak{C}(X, d_X) = (X, P)$. Consider $x, x' \in X$ and let $\delta = d_X(x, x')$. We can thus define a morphism $f(\Delta_2(\delta)) \longrightarrow X$ such that $f(p) = x$ and $f(q) = x'$. Because p and q are in the same block, by functorality x and x' are in the same block.

Scale Invariance

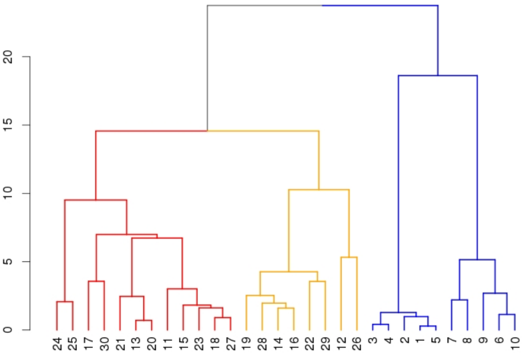
Case 2: $\mathfrak{C}(\Delta_2(1))$ has two clusters. Then by scale invariance $\mathfrak{C}(\Delta_2(\delta))$ has two clusters for all $\delta > 0$. Fix (X, d_X) and again write $\mathfrak{C}(X, d_X) = (X, P)$. We consider $x, x' \in X$, and partition X into subsets A and B such that $x \in A$ and $x' \in B$. Letting $\delta = \mathrm{sep}(X)$ we define $f : X \longrightarrow \Delta_2(\delta)$ by $f(y) = p$ if and only if $p \in A$. By functorality x and x' cannot be in the same block. $\square$

Clustering Functors on $\mathbb{M}^{iso}$

It's easy to classify clustering functors on $\mathbb{M}^{iso}$. Let $\mathcal{I}$ denote the set of isometry classes of finite metric spaces, and let (X_γ, d_γ) denote an element of class $\gamma \in \mathcal{I}$. As any clustering functor must act invariantly on self-symmetries of a given metric space, if we let G_γ denote the isometry group of (X_γ, d_γ) and Γ_γ the fixed points of the action of G_γ on $P(X)$, then a choice of clustering functor is equivalent to a choice of element in Γ_γ for all $\gamma \in \mathcal{I}$.

Hierarchical Clustering

We can also consider *Hierarchical Clustering*. Essentially, this is clustering where there exists a parameter which can be varied, increasing or decreasing the size of clusters. This kind of clustering can give us a better understanding of the data.



The Category $\mathbb{P}$

We wish to give a precise definition for Hierarchical clustering in terms of functors. The choices for input category are the same as before, but our output category will now consist of *persistent sets*. A persistent set $(X, \mathcal{O}_X)$ consists of a finite set X and a function $\mathcal{O}_X : \mathbb{R}_+ \longrightarrow \mathcal{P}(X)$ (partitions of X) such that $\mathcal{O}_X(r)$ is a refinement of $\mathcal{O}_X(r')$ whenever $r < r'$. We define the category $\mathbb{P}$ as having objects persistent sets and morphisms $f : (X, \mathcal{O}_X) \longrightarrow (Y, \mathcal{O}_Y)$ given by maps $f : X \longrightarrow Y$ such that $\mathcal{O}_X(r)$ refines $f^{-1}(\mathcal{O}_Y(r))$ for all $r \in \mathbb{R}_+$.

Hierarchical Clustering

A *Heirchical Clustering Functor* is a functor $\mathfrak{H} : \mathbb{M} \longrightarrow \mathbb{P}$ such that $\mathfrak{H} \circ \gamma = \alpha$, where $\gamma : \mathbb{P} \longrightarrow \mathsf{Set}$ is the forget functor.

We define the hierarchical Vietoris-Rips functor $\mathfrak{R} : \mathbb{M} \longrightarrow \mathbb{P}$ via $\mathfrak{R}(X, d_X) = (X, \mathcal{O}^{VR})$ where $\mathcal{O}^{VR}(r) = X/ \sim_r$.

Agglomerative Clustering Methods

We note that the $\mathfrak{R}$ is just agglomerative single-linkage clustering. There are other popular choices. Many can be constructed by starting with a linkage function l which measures distances between clusters. For $r > 0$ one can define the relationship $\sim_{l,r}$ on clusters as $B \sim_{l,r} B'$ if there exists $B = B_0, B_1, \dots B_k = B'$ with $l(B_i, B_{i+1}) \leq r$. We define a persistent set as follows: we define a sequence of partitions as $\mathcal{O}_1 = \{\{x\}\}_{x \in X}$ and $\mathcal{O}_{i+1} = \mathcal{O}_i / \sim_{l,r_i}$ with $r_i = \min\{l(B, B') : B, B' \in \mathcal{O}_i, B \neq B'\}$. We then define $\mathcal{O}^l(r) = \mathcal{O}_i$ where $i = \max\{r_i : r_i \leq r\}$.

Agglomerative Clustering Methods

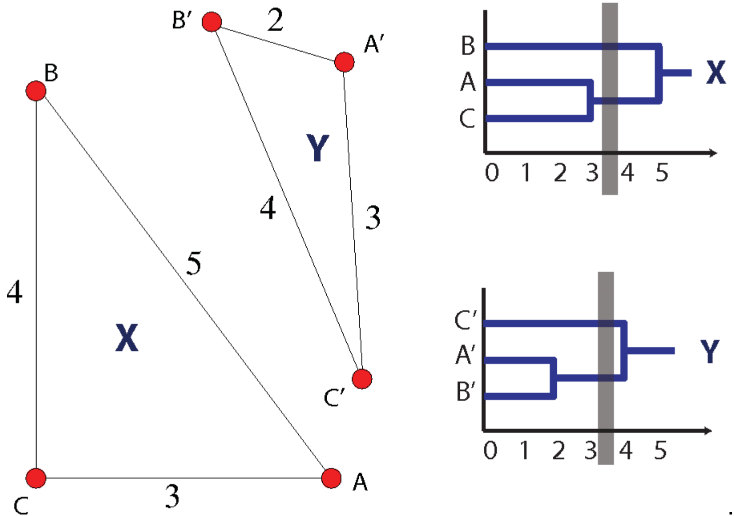
Some common linkage functions:

$$l_{sing}(B, B') = \min_{x \in B} \min_{x' \in B'} d_X(x, x')$$

$$l_{comp}(B, B') = \max_{x \in B} \max_{x' \in B'} d_X(x, x')$$

$$l_{avg}(B, B') = \frac{\sum_{x \in B} \sum_{x' \in B'} d_X(x, x')}{|B||B'|}$$

Agglomerative Clustering Methods



Average and complete linkage are not functorial over $\mathbb{M}^{inj}$!

A Uniqueness Theorem

Theorem

Let $\mathfrak{H} : \mathbb{M}^{gen} \longrightarrow \mathbb{P}$ *be a hierarchical clustering functor such that:*

- $\mathfrak{H}(\Delta_2(\delta)) = (\{p, q\}, \mathcal{O}_\Delta)$ *where* $\mathcal{O}_\Delta(r)$ *has two clusters for* $r < \delta$ *and one cluster for* $r \geq \delta$*.*
- *Writing* $\mathfrak{H}(X, d_X) = (X, \mathcal{O}^{\mathfrak{H}})$*, for* $r < \text{sep} X$*,* $\mathcal{O}^{\mathfrak{H}}(r)$ *consists of singletons.*

Then $\mathfrak{H} = \mathcal{R}$*.*

Proof of Uniqueness Theorem

Proof. We first show that $\mathcal{O}_X^{VR}(r)$ refines $\mathcal{O}_X^{\mathfrak{H}}(r)$ for all $r > 0$. It suffices to show that if $d_X(x, x') \leq r$ then x, x' lie in the same block in $\mathcal{O}_X^{VR}(r)$. If $d_X(x, x') \leq r$ then $p \mapsto x$, $q \mapsto x'$ defines a morphism $g : \Delta_2(r) \longrightarrow (X, d_X)$ in $\mathbb{M}^{gen}$. We thus get a morphism

$$\mathfrak{H}(g) : \mathfrak{H}(\Delta_2(r)) \longrightarrow \mathfrak{H}(X, d_X)$$

in $\mathbb{P}$. By assumption p, q lie in the same block of $\mathcal{O}_{\Delta_2(r)}(r)$, and thus by functorality x, x' lie in the same block of $\mathfrak{H}(X, d_X)$.

Proof of Uniqueness Theorem

Now we show that $\mathcal{O}_X^{\mathfrak{H}}(r)$ refines $\mathcal{O}_X^{VR}(r)$ for all $r > 0$. Consider the metric space (X_r, d_r) where the points are the equivalence classes of X under $\sim_r$ with

$$d_r(B, B') = \min_{x \in B} \min_{x' \in B'} d_X(x, x').$$

It's clear that $\mathrm{sep}(X_r) > r$ and thus $\mathcal{O}_{X_r}^{H}(r)$ consists precisely of the equivalence classes under $\sim_r$. We define $\pi_r : (X, d_X) \longrightarrow (X_r, d_r)$ which sends x to its equivalence class $[x]_r$. We thus get a map $\mathfrak{H}(\pi_r)$ which shows that $\mathcal{O}^{\mathfrak{H}}(r)$ refines $\mathcal{O}_{X_r}^{\mathfrak{H}}(r) = \mathcal{O}_X^{VR}(r)$. □

Kleinberg's Axioms

We note that not only is $\mathfrak{R}$ the only clustering functor satisfying the conditions of the previous theorem, it satisfies, in some sense, Kleinberg's 3 axioms.

- $\mathfrak{R}$ is scale invariant in the sense that the dendrogram of $\mathfrak{R}(X, d_X)$ is the same as for $\mathfrak{R}(X, \lambda \cdot d_X)$ for any $\lambda > 0$, up to a change of scale. More precisely, $\mathfrak{R} \circ \lambda \cdot (X, d_X) = \mathcal{O}^{VR}(r/\lambda)$.
- $\mathfrak{R}$ is *rich* in the sense that for any dendrogram $\mathcal{O}_X$, if $\epsilon_1, \ldots \epsilon_n$ are the transition points, then define $d_X(x, x')$ to be the smallest ϵ_i such that x and x' lie in the same cluster of $\mathcal{O}_X(\epsilon_i)$.
- As stated earlier, functorality is essentially a notion of consistency.

References

- Carlson, Gunnar and Facundo Memoli. *Classifying Clustering Schemes*. 2010
- Carlson, Gunnar and Facundo Memoli. *Persistent Clustering and a Theorem of J. Kleinberg*
- Luxburg, Ulrike von, Robert Williamson, Isabelle Guyon.*Clustering: Science or Art?*. 2009
- Riehl, Emily. *Functorality in Algebra and Topology*. 2017