Functorial cluster embedding

Steve Huntsman

FAST Labs / Cyber Technology
https://bit.ly/35DMQjr

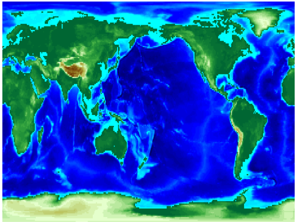
10 November 2019

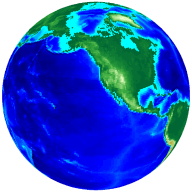
BAE SYSTEMS

Overview: TPE + functorial clustering = FCE

- Dimensionality reduction is a basic and ubiquitous approach for understanding high-dimensional data
  - Linear archetype: principal components analysis (PCA)
  - Most nonlinear dimensionality reduction (NLDR) techniques are *ad hoc*, even when motivated by or using theorems
- The NLDR technique of tree-preserving embedding (TPE) turns out to be functorial
- A category-theoretical classification of hierarchical clustering schemes gives a recipe for transforming TPE into essentially all functorial NLDR methods under the aegis of *functorial cluster embedding* (FCE)
  - Carlsson, G. and Mémoli, F. *JMLR* **11**, 1425 (2010); *Found. Comp. Math.* **13**, 221 (2013)
- Preceding two bullets essentially the only original material here

The quintessential NLDR example

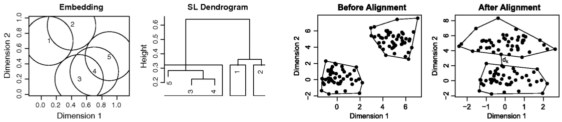




- 2D map results from applying NLDR to a globe surface in 3D
  - Different map projections suit varying purposes...
  - ...but tradeoffs are inevitable: e.g., topological information (a nontrivial homology class) must be lost unless the embedding has a point at infinity

Tree preserving embedding

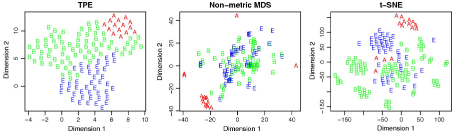
- For details see Shieh, A. D., *et al.* *PNAS* **108**, 16916 (2011)
- TPE preserves the single-linkage dendrogram
  - = hierarchical clustering of points resulting from merging cluster pairs with minimum nearest-neighbor distance
- How TPE does it:
  - Constrained optimization preserves the SL dendrogram
    - Acts directly on dissimilarities: no need for vector data
  - Infeasible in practice, but a good greedy approximation exists
    - Use an optimal rigid transformation of prior embedding instead of reembedding at each step
    - $O(n^3)$ runtime, typical for the class of NLDR algorithms



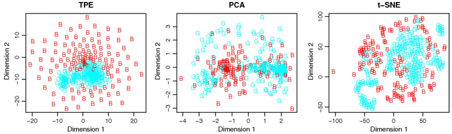
Images from Shieh *et al.*

TPE examples from Shieh *et al.*

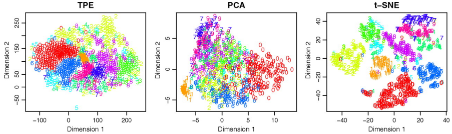
protein sequence dissimilarity (colors/labels for organism domains)



radar signals ($\in \mathbb{R}^{34}$, colors/labels for signal quality)



images of handwritten digits (colors/labels for digits themselves)



Relevant categories (see Carlsson and Mémoli)

- $\mathcal{M}^{\text{iso}} \subset \mathcal{M}^{\text{inj}} \subset \mathcal{M}^{\text{gen}}$: objects are finite metric spaces (X, d_X); morphisms are isometries/injective distance-nonincreasing maps
- $\mathcal{C}$ ("standard clustering algorithm outputs"): objects are (X, P_X), where P_X is a partition of X into clusters; morphisms are $f : X \to Y$ s.t. P_X refines $f^*(P_Y) := \{f^{-1}(B) : B \in P_Y\}$
- $\mathcal{P}$ ("hierarchical clustering algorithm outputs"): objects are *persistent sets* (X, θ_X) and morphisms are $f : (X, \theta_X) \to (Y, \theta_Y)$ s.t. $\theta_X(r) \leq f^*(\theta_Y(r))$ for all r
  - Here X is a finite set and θ_X is a map from $\mathbb{R}_{\geq 0}$ to the set of partitions of X s.t. i) $r \leq s \Rightarrow \theta_X(r) \leq \theta_X(s)$ and ii) for all $r \geq 0$ there exists $\epsilon > 0$ s.t. $\theta_X(r') = \theta_X(r)$ for all $r \leq r' \leq r + \epsilon$. A *dendrogram* is a persistent set (X, θ_X) s.t. $\theta_X(t)$ consists of a single cluster for some t

Relevant equivalence relations

- For $x, x' \in (X, d_X)$ and $r \geq 0$:
  - $x \sim_r x'$ iff there exists a sequence $x = x_0, x_1, \ldots, x_k = x'$ of points in X s.t. $d_X(x_j, x_{j+1}) \leq r$ for $0 \leq j \leq k-1$;
  - more generally, for any $m \in \mathbb{Z}_{\geq 0}$, an equivalence relation $\sim_r^m$ obtained by keeping equivalence classes under $\sim_r$ of cardinality $\geq m$ and associating any unaccounted-for points to singleton equivalence classes;
- For $B, B' \in P_X$, $R \geq 0$ and a *linkage function* ℓ defining the distance between clusters, $B \sim_{\ell,R} B'$ iff there exists a sequence $B = B_0, B_1, \ldots, B_k = B'$ of clusters in P_X s.t. $\ell(B_j, B_{j+1}) \leq R$ for $0 \leq j \leq k-1$.

Relevant functors

- *Standard clustering functor* $\mathfrak{C} : (\mathcal{M}^{iso}, \mathcal{M}^{inj}, \mathcal{M}^{gen}) \to \mathcal{C}$
  - Functoriality amounts to $(X, d_X) \xrightarrow{f} (Y, d_Y) \xrightarrow{\mathfrak{C}} (Y, P_Y) = (X, d_X) \xrightarrow{\mathfrak{C}} (X, P_X) \xrightarrow{\mathfrak{C}(f)} (Y, P_Y)$ w/ typical $\mathfrak{C}(f) = f$ in Set
- *Vietoris-Rips* or *single-linkage clustering functor* $\mathfrak{R}_r : \mathcal{M} \to \mathcal{C}$
  - $\mathfrak{R}_r(X, d_X) := (X, P_X(r))$, where $P_X(r)$ is the partition for $\sim_r$
  - $\mathfrak{R}_r(f : X \to Y)$ given by regarding f as a morphism from $(X, P_X(r))$ to $(Y, P_Y(r))$ in $\mathcal{C}$
- *Vietoris-Rips hierarchical clustering functor* $\mathfrak{R} : \mathcal{M}^{gen} \to \mathcal{P}$
  - $\mathfrak{R}(X, d_X) := (X, \theta_X)$ and where $\theta_X(r) = P_X(r)$ as above
  - $\mathfrak{R}(f : X \to Y)$ given by regarding f as a morphism from $(X, \theta_X(r))$ to $(Y, \theta_Y(r))$ in $\mathcal{P}$

BAE SYSTEMS

Representable/excisive standard clustering functors

- More general class of standard clustering functors than $\mathfrak{R}_r$
  - Defined in terms of a family Ω of finite metric spaces
- $\mathfrak{C}^\Omega : \mathcal{M} \to \mathcal{C}$ is given by $\mathfrak{C}^\Omega(X, d_X) := (X, P_X)$
  - Here x and x' belong to the same cluster of P_X iff there exists a sequence $x = x_0, x_1, \ldots, x_k = x'$ of points in the cluster, along with $\{\omega_j\}_{j=1}^k \subseteq \Omega$, $(\alpha_j, \beta_j) \in \omega_j^2$, and $f_j \in \hom_{\mathcal{M}}(\omega_j, X)$ for $0 \leq j \leq k-1$ s.t. $f_j(\alpha_j) = x_{j-1}$ and $f_j(\beta_j) = x_j$.
  - Example: $\mathfrak{R}_r = \mathfrak{C}^{\{\Delta_2(r)\}}$, where $\Delta_m(r)$ denotes the metric space with m points each at distance r from each other
- Theorem: $|\Omega| < \infty \Rightarrow \mathfrak{C}^\Omega = \mathfrak{R}_1 \circ \mathfrak{I}^\Omega$
  - $\mathfrak{I}^\Omega$ is a metric-changing endofunctor with a specific formula
- Uniqueness results also highlight the special nature of $\mathfrak{R}_r$

The metric-changing endofunctor

- $\mathfrak{I}^\Omega(X, d_X) := (X, \mathcal{U}(W_X^\Omega))$
- *Maximal subdominant ultrametric* $\mathcal{U}(W_X)$
    - W/r/t symmetric $W_X : X^2 \to \mathbb{R}_{\geq 0}$ w/ $W_X(x,x) \equiv 0$
    - $\mathcal{U}(W_X)(x,x') := \min \{\max_{x=x_0,x_1,\dots,x_k=x'} W_X(x_j, x_{j+1})\}$
    - I.e., the maximal hop in a minimal path between points
    - Algorithm provided in §VI.C of Rammal, Toulouse, and Virasoro, *Rev. Mod. Phys.* **58**, 765 (1986)
- $W_X^\Omega(x,x') := 0$ if $x = x'$, otherwise equals $\inf \{\lambda > 0 : \exists \omega \in \Omega, \phi \in \hom_{\mathcal{M}}(\lambda \cdot \omega, X) \text{ s.t. } \{x,x'\} \subset \phi(\lambda \cdot \omega)\}$
- Example: for $\Omega = \{\Delta_m(\delta)\}$ we have $W_X^\Omega(x,x') = \inf \{\lambda > 0 : \exists X_m \subset X \text{ s.t. } |X_m| = m \wedge \{x,x'\} \subset X_m \wedge d_X|_{X_m} \leq \lambda\delta\}$
    - Find a min-diameter subset with m elements including x and x'
    - Generally have to use heuristics

Remarks on density proxies and hierarchical clustering

- Density estimates in high dimensions will generally be poor
  - Functoriality is a more reasonable desideratum for clustering than density recognition
  - This point of view supports "functorial NLDR" and simple Ω
- Theorem: $\mathfrak{R}$ is the unique hierarchical clustering functor on $\mathcal{M}^{gen}$ that satisfies a few mild/natural restrictions
  - More options on $\mathcal{M}^{inj}$
    - Let $\theta_X^m(r)$ be the partition of (X, d_X) w/r/t $\sim_r^m$. Now $\mathfrak{H}^m : \mathcal{M}^{inj} \to \mathcal{P}$ defined by $\mathfrak{H}^m(X, d_X) := (X, \theta_X^m)$ (and the trivial action on maps) works; clustering amounts to treating small numbers of co-located "outliers" as singletons
    - A particularly useful class of hierarchical clustering functors is furnished by taking $\mathfrak{R}^\Omega := \mathfrak{R} \circ \mathfrak{I}^\Omega$, e.g., hierarchical-functorial analogue of DBSCAN

Functorial cluster embedding

- Generalization from TPE to FCE is significant yet easy
- Given a hierarchical clustering functor $\mathfrak{R}^{\Omega} : \mathcal{M}^{inj} \to \mathcal{P}$, to elegantly embed (X, d_X) in some $\mathbb{R}^n$ we merely need to:
    - apply $\mathfrak{I}^{\Omega}$ to (X, d_X);
    - perform TPE
- FCE preserves $\mathfrak{R}^{\Omega}$ since TPE preserves $\mathfrak{R}$
    - I.e., FCE simply amounts to the observation that TPE is essentially functorial over $\mathcal{M}^{gen}$ along with the application of the endofunctor $\mathfrak{I}^{\Omega}$
- Example: $\Omega = \{\Delta_m(\delta)\}$ leads to a hierarchical-functorial analogue of "DBSCAN-tree preserving embedding" likely to enhance the utility of TPE

Implementing FCE

- A practical implementation of FCE requires:
  1) An algorithm taking the original metric d_X as input and producing a symmetric function of the form W^Ω as output;
  2) An algorithm for computing the subdominant ultrametric;
  3) An implementation of TPE itself
- Items 2 & 3 are straightforward/available, though existing implementation of TPE restricts embedding to $\mathbb{R}^2$
- Item 1 will generally be **NP**-hard for a nontrivial choice of Ω
  - Constrain Ω
  - Accept approximate solutions (already doing this for TPE)

Implementation notes for $\Omega = \{\Delta_m(\delta)\}$

- For $m = 3$ we can avoid any bottleneck:
  - $W_X^\Omega(x, x') = \inf \left\{ \lambda > 0 : \exists x'' \in X \text{ s.t. } d_X|_{\{x,x',x''\}} \leq \lambda\delta \right\}$ takes $O(n^3)$ steps–same as subdominant ultrametric and TPE
- For $m > 3$, let $H_k(x)$ denote the k points closest to x, including x itself, and approximate $W^\Omega(x, x')$ for $m/2 \geq k = \Theta(m)$ by restricting consideration from X to $H_k(x) \cup H_k(x')$ in formation of m-element min-diameter sets
  - Helpful to precompute a hash table of sets of indices corresponding to m-element subsets of $H_k(x) \cup H_k(x')$
- Can employ greedy approximations, particularly for $X \subset \mathbb{R}^N$
- Some other more esoteric tactics might be considered

Conclusion

- Provides principled basis for developing practical instantiations: focus on approximation of nice algorithms instead of efficient but ad hoc constructions
- Category theory can help us recognize (what) a good thing (is) when we see it...
- ...and we can miss good things by not paying attention to the categorical context

Thanks!

steve.huntsman@baesystems.com
https://bit.ly/35DMQjr

BAE SYSTEMS