Optimal Transformations of High-dimensional Functional Data for Clustering Methods

Hanchao Zhang

Division of Biostatistics, Department of Population Health,
Grossman School of Medicine, New York University

August 9, 2022

NEW YORK UNIVERSITY

NYU Grossman School of Medicine

Acknowledgements

This talk presents joint work with:
Thaddeus Tarpey (Grossman School of Medicine, New York University)

This work is supported by grant:
R01 MH099003 from the National Institute of Mental Health (NIMH)

Outline for section 1

The Diagnostic and Statistical Manual of Mental Disorders (DSM-5)

Major Derpression (partial criteria)

- Depressed mood most of the day, nearly every day, as indicated by either subjective report (e.g. feels sad, empty, hopeless) or observation made by others (e.g., appears tearful). (**Note**: In children and adolescents, can be irritable mood.)
- Markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day (as indicated by either subjective account or observation).
- Significant weight loss when not dieting or weight gain (e.g., a change of more than 5% of body weight in a month), or decrease or increase in appetite nearly every day. (**Note**: In children, consider failure to make expected weight gain.)
- Insomnia or hypersomnia nearly every day.
- Fatigue or loss of energy nearly every
  . . . .

Bipolar (partial criteria)

- Depressed mood most of the day, nearly every day, as indicated by either subjective report (e.g. feels sad, empty, hopeless) or observation made by others (e.g., appears tearful). (**Note**: In children and adolescents, can be irritable mood.)
- Markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day (as indicated by either subjective account or observation).
- Significant weight loss when not dieting or weight gain (e.g., a change of more than 5% of body weight in a month), or decrease or increase in appetite nearly every day. (**Note**: In children, consider failure to make expected weight gain.)
- Insomnia or hypersomnia nearly every day.
- Fatigue or loss of energy nearly every
  . . . .

[1]

[1] Exclusion of overlapping symptoms in DSM-5 mixed features specifier: heuristic diagnostic and treatment implications

The Diagnostic and Statistical Manual of Mental Disorders (DSM-5)

Major Derpression (partial criteria)

- ......
- Depressed mood most of the day, nearly every day, as indicated by either subjective report (e.g. feels sad, empty, hopeless) or observation made by others (e.g., appears tearful). (**Note**: In children and adolescents, can be irritable mood.)
- Markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day (as indicated by either subjective account or observation).
- ......

Bipolar (partial criteria)

- ......
- Depressed mood most of the day, nearly every day, as indicated by either subjective report (e.g. feels sad, empty, hopeless) or observation made by others (e.g., appears tearful). (**Note**: In children and adolescents, can be irritable mood.)
- Markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day (as indicated by either subjective account or observation).
- ......

Goals

- Predict the current diagnosis of the patient.
- Subtype the patient's into finer groups.

Outline for section 2

Introduction to Functional Data

Functional Data

$\boldsymbol{y}_i(t), i = 1, \ldots, n, t \in T$, typically a compact real interval

$$\boldsymbol{y}_i(t) = \boldsymbol{\theta}'(t)\boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i(t) = \sum_{j=1}^{\infty} \beta_{ij}\theta_j(t) + \epsilon_i(t)$$

$\boldsymbol{\theta} = (\theta_1(t), \ldots, \theta_p(t), \ldots)'$ is a vector basis observations represented by basis functions

$\boldsymbol{\beta}_i = (\beta_{1i}, \ldots, \beta_{ip}, \ldots)'$ is a vector of regression coefficients

- MRI and fMRI data [Chen, Reiss, and Tarpey, 2014]
- EEG data of Brain [Jiang, Petkova, Tarpey, and Ogden, 2017]
- Functional data can be seen as trajectories and expressed using basis representations.

Clustering Functional Data

Using Functional Coefficients as the Clustering Data

- $\boldsymbol{y}_i(t), i = 1, \ldots, n, t \in T$, typically a compact real interval
- $\boldsymbol{Y} = (y_1(t), y_2(t), \ldots, y_n(t))'$
- $\beta_i = (\beta_{1i}, \ldots, \beta_{ip}, \ldots)'$ is a vector of regression coefficients
- $\boldsymbol{B} = (\beta_1, \beta_2, \ldots, \beta_n)'$ is the matrix of the basis coefficients

The dimension of $\boldsymbol{B}$ depends on the dimension of the basis function we've chosen in the previous step.

Clusering Functional Data [Tarpey and Kinateder, 2003] [a]

[a] in this talk, we perform clustering based on basis coefficients

- perform clustering algorithm based on raw data $\boldsymbol{Y}$
- perform clustering algorithm based on basis coefficients $\boldsymbol{B}$

Clustering Methods K-Means and Generalized K-Means

Data

Define $\boldsymbol{x} := \boldsymbol{B}$, instead of using raw data $\boldsymbol{Y}$, use the basis coefficients $\boldsymbol{B}$

K-Means

$$\left.\begin{array}{lll} \text{minimize}_{\mathcal{C}} \; g_n(\mathcal{C}) & = \frac{1}{n}\sum_{i=1}^{m}\sum_{k\in\mathcal{C}_i}||x_k - \bar{x}_{\mathcal{C}_i}||^2 \\ \text{maximize}_{\mathcal{C}} \; h_n(\mathcal{C}) & = \sum_{i=m}\frac{|\mathcal{C}_i|}{n}\cdot||\bar{x}_{\mathcal{C}_i}||^2 \end{array}\right\} \text{equivalent}$$

Conveixty-Based Clustering [Bock, 2004]

$$\text{maximize}_{\mathcal{C}} \quad \tilde{h}_n(\mathcal{C}) = \sum_{i=m}\frac{|\mathcal{C}_i|}{n}\cdot\phi(\bar{x}_{\mathcal{C}_i})$$

Where ϕ is any arbitrary convex function. ($\phi = ||\cdot||^2$ for K-means)

2

[2] the following data represented by $\boldsymbol{x}$ are the basis coefficients $\boldsymbol{B}$

Convexity-Based Clustering in a Continuous Format

Continuous Format

Consider a random variable X in $\mathbb{R}^p$ with some probability distribution P, and look for m partitions $\mathcal{B} = (\mathcal{B}_1, \mathcal{B}_2, \dots, \mathcal{B}_m)$ of the entire space $\mathbb{R}^p$

$$\textbf{maximize}_{\mathcal{B}} \ H(\mathcal{B}) := \sum_{j=1}^{m} P(\mathcal{B}_i) \cdot \phi(E[X|X \in \mathcal{B}_i])$$

- $X \in \mathbb{R}^p$ random variable
- P is the distribution of random variable X
- $\mathcal{B} = (B_1, B_2, \dots, B_m)$ is the partition of the entire space $\mathbb{R}^p$

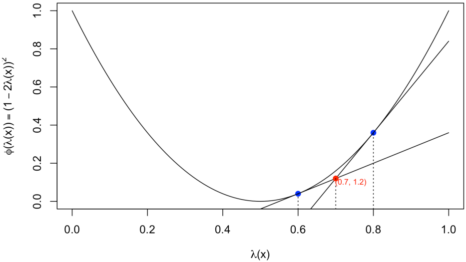
Continuous Format with Two Systems

$$G(\mathcal{B}, \mathcal{Z}) := \sum_{i=1}^{m} \int_{\mathcal{B}_i} \left[\phi(x) - t(x; z_i)\right] dP(x) = E[\phi(X)] - E[p(X; \mathcal{B}, \mathcal{Z})]$$

Visualization of The Partition $\mathcal{B}$ and Center $\mathcal{Z}$

Continuous Format with Two Systems

$$G(\mathcal{B}, \mathcal{Z}) := \sum_{i=1}^{m} \int_{\mathcal{B}_i} \left[\phi(x) - t(x; z_i)\right] dP(x) = E[\phi(X)] - E[p(X; \mathcal{B}, \mathcal{Z})]$$



- $t(x; z_i)$ is the support tangent plane
- ● Boundary of the partition $\mathcal{B}$
- ● Centers system $\mathcal{Z}$ of partition $\mathcal{B}$ defined by support hyperplane

Maximum Support Plane Algorithm (MSP)

Continuous Format with Two Systems

$$G(\mathcal{B}, \mathcal{Z}) := \sum_{i=1}^{m} \int_{\mathcal{B}_i} \left[\phi(x) - t(x; z_i)\right] dP(x) = E[\phi(X)] - E[p(X; \mathcal{B}, \mathcal{Z})]$$

Algorithm 1 **Maximum Support Plane Algorithm (MSP)**

1: $t = 0$ start with an initial system $\mathcal{Z}^{(0)} = (z_1^{(0)}, \ldots, z_m^{(0)})$, m distinct support points from $\mathbb{R}^p$
2: **while** $\mathcal{Z}^{(t)} \neq \mathcal{Z}^{(t-1)}$ **do**
3: $t = t + 1$
4: determine a m-partition $\mathcal{B}^{(t+1)} = \arg\min_{\mathcal{B}} G(\mathcal{B}, \mathcal{Z}^{(t)})$
5: determine the support points $\mathcal{Z}^{(t+1)} = \arg\min_{\mathcal{Z}} G(\mathcal{B}^{(t+1)}, \mathcal{Z})$
6: **end while**

3

[3] Computationally intensive in high-dimension using MCMC for support hyperplane

Outline for section 3

Choosing Convex Function $\phi(\cdot)$

Bayesian Decision Rule (2 classes)

$$f(\boldsymbol{x}) = \pi_1 f_1(\boldsymbol{x}) + \pi_2 f_2(\boldsymbol{x})$$

where $f_1(x)$, $f_2(x)$ are the probability density functions, and π_1, π_2 are the prior probability of the two subpopulations.

Posterior Probability

$$\lambda(\boldsymbol{x}) = \frac{\pi_2 f_2(\boldsymbol{x})}{\pi_1 f_1(\boldsymbol{x}) + \pi_2 f_2(\boldsymbol{x})}$$

$\lambda(x)$ is the posterior probability of classifying x to the second population.

Convex Function $\phi(x)$

$$\phi(\boldsymbol{x}) = (1 - 2\lambda(\boldsymbol{x}))^2 = \left(\frac{\pi_1 f_1(x) - \pi_2 f_2(x)}{\pi_1 f_1(x) + \pi_2 f_2(x)}\right)^2$$

The $\phi(x)$ function can be viewed as the “purity” of the classifier.

Fast MSP Algorithm

Skip Some Derivations .....

Algorithm 2 **Fast Maximum Support Plane Algorithm (Fast MSP)**

1: calculate posterior probability $\lambda(\boldsymbol{x}) = \frac{\pi_2 f_2(\boldsymbol{x})}{\pi_1 f_1(\boldsymbol{x}) + \pi_2 f_2(\boldsymbol{x})}$

2: perform K-means on $\lambda(\boldsymbol{x})$

Take Home Message of the Fast MSP Algorithm

- Reduce the high-dimensional data into a probability scalar
- Perform sub-typing based on the probability scalar
- Utilize baseline information obtain from basis coefficients

Estimating Density Functions of Sub-populations

Independent Component Factorization

Let $\boldsymbol{X}^c$ be the centered coefficients representation of raw data.

$$\boldsymbol{X}^C_{n\times p} = \boldsymbol{S}_{n\times p}\boldsymbol{W}^{-1}_{p\times p}$$

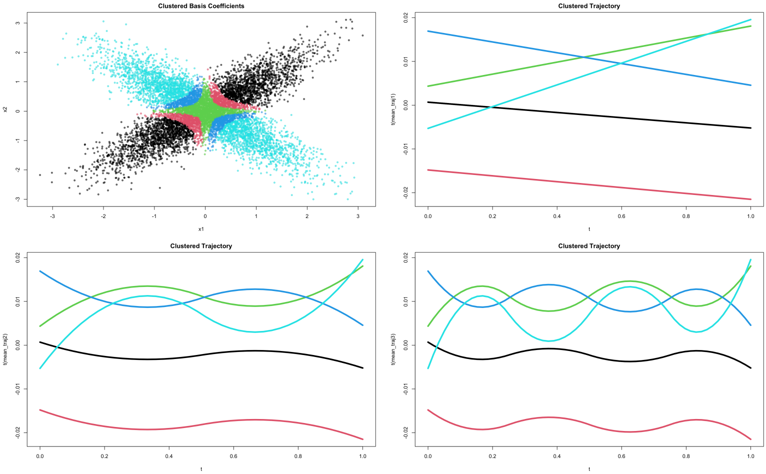
- $\boldsymbol{X}^c$ is the centered data matrix $\boldsymbol{X}$
- $\boldsymbol{W}$ is a whitening matrix
- $\boldsymbol{S}$ contains the independent components

Kernel Density for $f_j(s)$

Estimating $f_j(\boldsymbol{x})$ using $f_j(s) = \prod_{\ell=1}^{p} f_{j\ell}(s_\ell)$, $j = 1, 2$, and $\ell = 1, 2, \ldots p$

Convexity-Based Clustering Results

Results and Embed to Higher Dimensions



Outline for section 4

Conclusion

Summary

- Represent the functional data using basis coefficients
- Clustering the functional data based on the basis coefficients
- Sub-typing the functional data into finer groups
- Capture the rich functional information

Referenes

Gin S Malhi, Yulisha Byrow, Tim Outhred, and Kristina Fritz. Exclusion of overlapping symptoms in dsm-5 mixed features specifier: heuristic diagnostic and treatment implications. *CNS spectrums*, 22(2):126–133, 2017.

Huaihou Chen, Philip T Reiss, and Thaddeus Tarpey. Optimally weighted l2 distance for functional data. *Biometrics*, 70(3):516–525, 2014.

Bei Jiang, Eva Petkova, Thaddeus Tarpey, and R Todd Ogden. Latent class modeling using matrix covariates with application to identifying early placebo responders based on eeg signals. *The annals of applied statistics*, 11(3):1513, 2017.

Thaddeus Tarpey and Kimberly KJ Kinateder. Clustering functional data. *Journal of classification*, 20(1), 2003.

Hans-Hermann Bock. Convexity-based clustering criteria: theory, algorithms, and applications in statistics. *Statistical Methods and Applications*, 12(3):293–317, 2004.

Thank You