HOW TO REGISTER TWO CURVES AS A BAYESIAN?

H.LUO

ABSTRACT. The term "register" means alignment in some sense, this subject is studied in the background of "standardization two curves" to make them somehow comparable.

(1) Ways of registering curves
Registration is to align two curves via some allowable transformations which usually form a group G.[1] Therefore we are doing some sort of standardization of two curves C_1, C_2 into a single "mean" curve $\bar{C}$ and try to compare them using the distance between $C_i, \bar{C}$. Ordinary Procrustes analysis is adopted in this paper[2]. The concern is mostly on how to standardize; if the concern is on how to compare, then we have already encountered elastic metric in [5]. Here the term "curve" is rather arbitrarily used, when dimension is one we call them functions; dimension is two we call them planar curves and so on.

(2) Space of concern
   (a) Different spaces.
      (i) Original space O. This is the space where the curves originally live in. Landmarks characterization is [2] O is chosen to be not the curves but k landmark features lying in $\mathbb{R}^m$, therefore $O = \mathbb{R}^{mk} - 0$. This can be regarded as a Grassmannian as indicated by Michor-Mumuford framework. [3]
      (ii) Ambient space S. This is the space where the domain/co-domain of the curves is regularized. For example, we usually make the domain to be $[0, 1]$ or $\mathbb{S}^1$ for convenience.
      (iii) Quotient space $Q = S/G$. This is the space where the objects are equivalent up to transformations contained in G.
   (b) Different metrics among curves.
   Here again we have to be careful about what kinds of metric is well-defined on the quotient spaces. In Michor-Mumford framework this requirement is described as the requirement that the immersion is free OR the G acts on the S transitively.
      (i) Metric on S.

Date: 07-19-2016.

[1] Actually a Lie group, I found Michor-Mumuford idea of using Lie group as a computational tool more and more attractive.

[2] In short, transform-impose scheme

[3] They are not analyzing an infinite dimensional space so that is why they can use $SO(m)$. However, the space of curves lying in $\mathbb{R}^m$ is still infinite dimensional unless we replace the curves by their landmark coordinates.

(ii) Induced metric on Q. $d_Q([C_1],[C_2]) := inf_{\gamma \in G} d_S(C_1, C_2 \circ \gamma)$ (parallel orbit property)

(iii) Elastic metric on Q. We must use the SRVF representation of curves in order to use elastic metric to induce an appropriate metric on the quotient space. The benefit is that elastic metric is invariant under reparametrization and we do not have to consider any G containing a diffeomorphism group, which is hard to deal with.

(A) $dim = 1$ The Fisher-Rao metric. Standard curve

$$\hat{\mu_Q} = arginf_{\mu \in Q} \sum_{i=1}^{n} inf_{\gamma_i \in G} d_Q([\mu], [\sqrt{\gamma_i'}(C_i \circ \gamma_i)])$$

where γ_i is the element in G that makes the SRVF representation of sample C_i deviates from $[\mu]$ as little as possible. Summing up these minimized distances and minimize over Q we yield the desired Frechet mean.

(B) $dim = 2$ The Younes elastic metric. [5]

$$\hat{\mu_Q} = arginf_{\mu \in Q} \sum_{i=1}^{n} inf_{\gamma_i \in G, \Gamma_i \in SO(m)} d_Q([\mu], [\sqrt{\gamma_i'}(C_i \circ \gamma_i)\Gamma_i])$$

where γ_i is the element in G that makes the SRVF representation of sample C_i deviates from $[\mu]$ as little as possible. Summing up these minimized distances and minimize over Q we yield the desired Frechet mean.

(c) Different means.

(i) Mean on S. $\mu_A = arginf_{\nu \in S} \int_S d_S(C, \nu) h(C) dC$ where $h(C)$ is the probability density indicating the probability of occurrence of curve C.[4]

(ii) Mean on Q. $\mu_Q = arginf_{[\nu] \in Q} \int_Q d_Q([C], [\nu])[h]([C]) d[C]$ where $[h]([C])$ is the probability density indicating the probability of occurrence of equivalent class $[C] \in Q$ whose representative is curve C. Explicitly speaking, we can induce $[h]$ using $[h] := \int_G h(C \circ \gamma) d\gamma$ for a probability already defined on S as long as the group action of G is transitive.
This construction is much like the Polya's urn scheme pooling all the elements in the same orbit of G into one urn.

(A) If a global infimum exists, then it is called a Frechet mean.

(B) If only local infimums exists, then it is called a Karcher mean.

(d) Different characterizations of γ.
The objective is to find a γ_i for each sample curve C_i such that $\sqrt{\gamma_i'}(C_i \circ \gamma_i) = \bar{C}$, $\bar{C}$ is the standard of the sample $\{C_1, \cdots C_n\}$. However, when the strict equality is not attainable, we must choose γ by minimize the sum of distances from C to $\bar{C}$ to yield such a curve. The standard curve formula is given for elastic metric above.

[4] If they do not adopt the landmark coordinates then $h(C)$ will be defined on an infinite dimensional space and this mean is not even tractable.

(i) B-spline basis. Represent the γ using a B-spline basis and then we can use a vector to represent a $\gamma \in G$.
(ii) Hallmarks. Represent the γ using $p_i = \gamma(t_i) - \gamma(t_{i-1})$ for a group of sampling points $\{t_1, \cdots t_M\}$.
(iii) (My proposal) Represent γ using its normal field.

(3) Choice of priors[5]
$\bar{C}$(standard curve)$\rightarrow \bar{C} \cdot Gaussian\, Noise$(oscillated curve)$\rightarrow \kappa^{\frac{p}{2}} \cdot \bar{C} \cdot Gaussian\, Noise \circ$ γ(unaligned curve)
  (a) Gaussian process for "noise" from the standard $\bar{C}$.
  (b) Dirichlet prior for $\gamma \in G$.[6]
  (c) Gamma prior for concentration number κ.

(4) Comparison of clustering effect
OPA $Choose\, \bar{C} \rightarrow Transform{-}Impose C_1\, to\, \bar{C} \rightarrow Transform{-}Impose C_2\, to\, \bar{C}$
GPA $Choose\, \bar{C} \rightarrow Transform - Impose\, C_1\, to\, \bar{C} \rightarrow Adjust\, \bar{C}\, to\, \bar{C_{new}} \rightarrow Transform - Impose\, C_2\, to\, \bar{C_{new}}$
  (a) Training shapes are clustered using Ordinary Procrustes analysis, testing shapes classified by Ordinary Procrustes analysis to minimize the Euclidean distance between landmarks coordinates.
  (b) Training shapes are clustered using Bayesian registration, testing shapes classified by Bayesian registration to minimize the elastic distance between landmark coordinates.(B-SRVF)
  (c) Training shapes are clustered using Generalized Procrustes analysis, testing shapes classified by SVM. (SRVF) [7]
  (d) **Training shapes are clustered using Bayesian registration, testing shapes classified by SVM. (BEST, time-consuming)**

(5) My comments
  (a) They are not using real nonparametric priors, which is a natural choice for the space of curves, they take m "hall marks" (equally spaced) among the curve and to give it a prior of finite dimension. This can be cured by using the Polya tree process over the normal field. i.e. We can directly assign a nonparametric prior on the normal field of a curve. My thinking here is approximately close to the "tail-free" prior measure used by [3] yet their use is still based on a B-spline basis and thus not really nonparametric.
  (b) They are using a "flawed" prior. The problem of alignment is very closely related to the location problem associated with the Dirichlet prior pointed out as a potential pathological example when the data is not lying in the Kullback-Leibler support of the prior. [4] The Bernstein-von Mises theorem only asserts pointwise consistency which

[5] There are two places where infinite dimension is the nature. One is the space of curves in $\mathbb{R}^m$, which is eliminated by using landmark coordinates; the other is the group G as a Lie group is of infinite dimension in nature yet it is eliminated by using the Hallmark technique we mentioned above. I do not think it is a good way of handling this infinite dimension.

[6] They must be very careful when they claimed that Simulated tempering can eliminate the multimodality, I guess they will still encounter trouble if the data is highly diffused and the Dirichlet prior has a heavy-tail base measure α. See [4].

[7] In short, like k-centroids we divide multiple shapes into a few groups with one standard shape(updated while clustered) for each group using transform-impose scheme. [2]

does not mean anything if we are analyzing in the space of curves. See my "CCBN" notes, Chap.3.

(c) The computational time is awful even if it is an algorithm based on simulation. The improvement of B-SRVF(Bayesian SRVF representation) over SRVF is not very big and according to their supplementary material the computational time is simply awful. We can even use an MLE via the first few terms of Laplace expansion of posterior if only approximate preciseness is needed.

References

[1] Cheng, Wen, Ian L. Dryden, and Xianzheng Huang. "Bayesian registration of functions and curves." Bayesian Analysis 11.2 (2016): 447-475.

[2] Kendall, David G. "A survey of the statistical theory of shape." Statistical Science (1989): 87-99.

[3] Claeskens, Gerda, Bernard W. Silverman, and Leen Slaets. "A multiresolution approach to time warping achieved by a Bayesian prior–posterior transfer fitting strategy." Journal of the Royal Statistical Society: Series B (Statistical Methodology) 72.5 (2010): 673-694.

[4] Diaconis, Persi, and David Freedman. "On the consistency of Bayes estimates." The Annals of Statistics (1986): 1-26.

[5] Kurtek, Sebastian, et al. "Statistical modeling of curves using shapes and related features." Journal of the American Statistical Association 107.499 (2012): 1152-1165.