Supervised Classification for Functional Data Using Multivariate Functional Depth

Chong Ma, David B. Hitchcock   University of South Carolina, Department of Statistics

Statistical Depth Function

Statistical depth function serves as a tool providing a center-outward ordering of points in $\mathbb{R}^d$, i.e., $D(\cdot;\cdot): \mathbb{R}^d \times \mathfrak{F} \to \mathbb{R}^+ \cup \{0\}$. It should ideally satisfy properties:

1. Affine invariance.
2. Maximality at center.
3. Monotonicity relative to deepest point.
4. Vanishing at infinity.

Depth Distribution

Assume $\{X(t), t \in \mathcal{I}\}$ is a stochastic function up to p^{th} differentiable. Denote $\mathbf{X} = \{\mathbf{X}(t), t \in \mathcal{I}\}$ where $\mathbf{X}(t) = (x^{(0)}(t), x^{(1)}(t), \ldots, x^{(p)}(t))$. The depth function of $\mathbf{X}$ w.r.t $\mathbf{Y} \sim \mathbb{F}_{\mathbf{Y}}$ is a univariate variable related to $\mathbf{X}$, i.e.,

$$D(\mathbf{X}; \mathbb{F}_{\mathbf{Y}}) \sim G$$

where $G: [0, \infty] \to [0, 1]$ is the cdf of the depth value $D(\mathbf{X}; \mathbb{F}_{\mathbf{Y}})$.

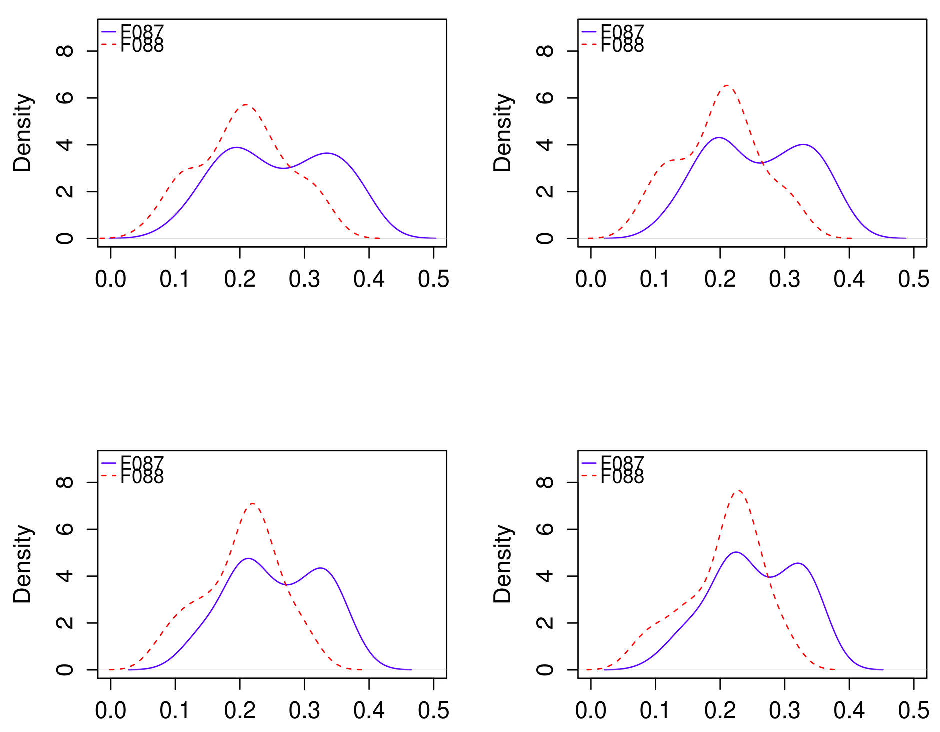

Figure 1: Depth Density Plot using 1st, 2nd, 3rd and 4th derivatives successively.

12 Blue Acrylic Fibers

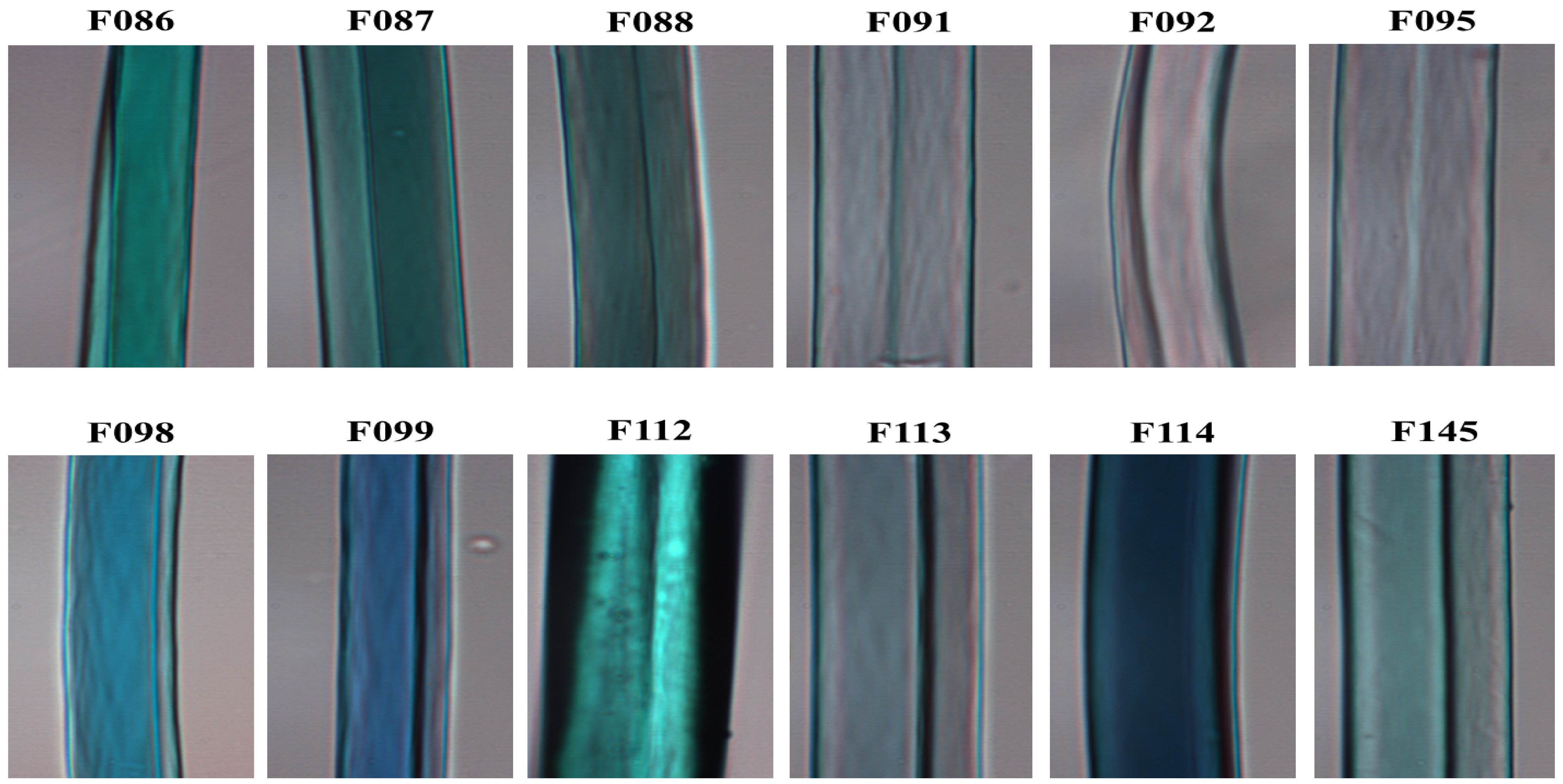

Figure 2: Microscopic images of 12 blue acrylic fibers under 40 × magnification

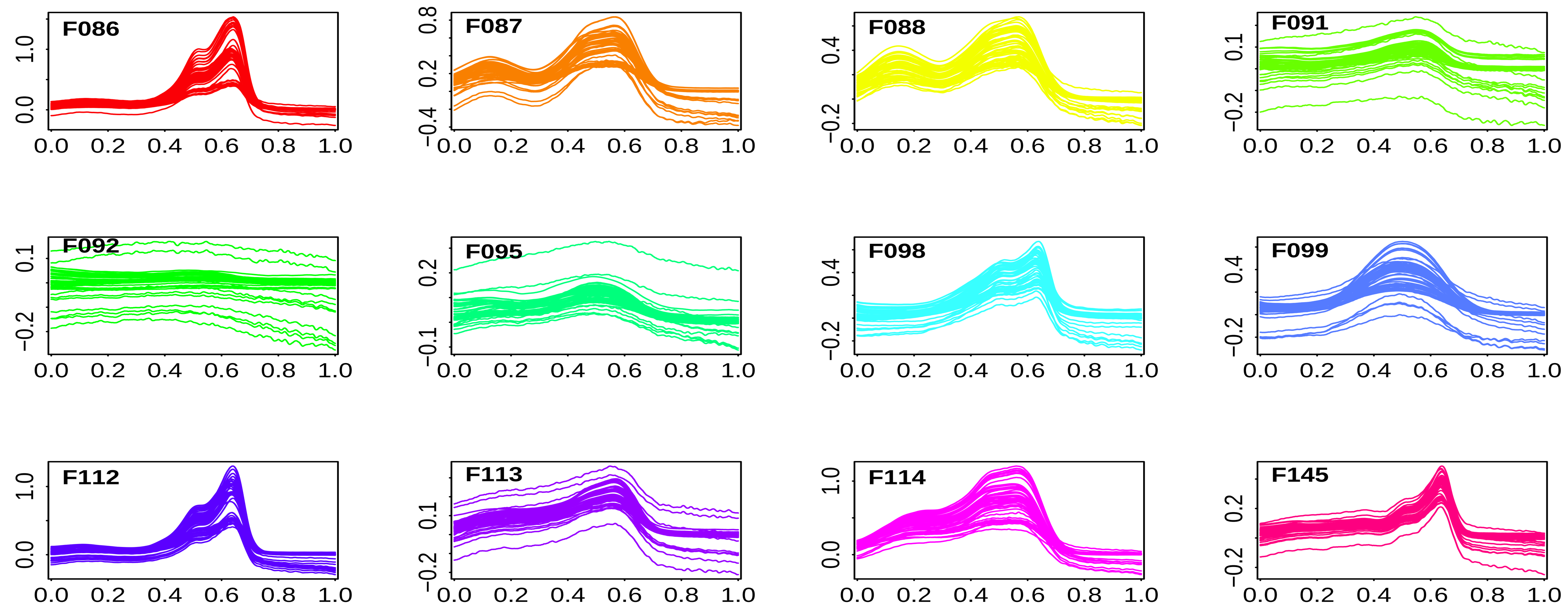

Figure 3: UV-visible absorbance spectra of 12 blue acrylic fibers

Bayesian Classifier

Consider two groups of curves

$$\mathbf{X}_1, \ldots, \mathbf{X}_n \overset{i.i.d}{\sim} \mathbb{F}_{\mathbf{X}}$$

and

$$\mathbf{Y}_1, \ldots, \mathbf{Y}_m \overset{i.i.d}{\sim} \mathbb{F}_{\mathbf{Y}}$$

To classify a new curve $\mathbf{Z}$ with the unknown class membership, we may test the hypothesis test

$$H_0: \mathbf{Z} \sim \mathbb{F}_{\mathbf{X}} \text{ vs. } H_a: \mathbf{Z} \sim \mathbb{F}_{\mathbf{Y}}$$

Let $D(\mathbf{Z}) = (D(\mathbf{Z}; \mathbb{F}_{\mathbf{X}}), D(\mathbf{Z}; \mathbb{F}_{\mathbf{Y}}))'$, and Γ be the rejection region. Assume H_0 is true with a priori probability π_0 and π_1 for H_a such that $\pi_0 + \pi_1 = 1$. The optimal Bayesian classifier is

$$\min_{\Gamma}\{FDR(\Gamma) + 1 - TPV(\Gamma)\}$$

where

$$FDR(\Gamma) = \frac{\pi_0 \Pr(D(\mathbf{Z}) \in \Gamma | \mathbf{Z} \sim \mathbb{F}_{\mathbf{X}})}{\Pr(D(\mathbf{Z}) \in \Gamma)}$$

and

$$TPV(\Gamma) = \frac{\pi_0 \Pr(D(\mathbf{Z}) \notin \Gamma | \mathbf{Z} \sim \mathbb{F}_{\mathbf{X}})}{\Pr(D(\mathbf{Z}) \notin \Gamma)}$$

Application Data

The UV-visible absorbance spectra of 12 blue acrylic fibers were examined at five separate locations including three academic research laboratories and two forensic laboratories. Each laboratory measures 10 replicate spectra of each fiber at different locations along the length of the fiber. The spectra were in a fine-grid of 1,175 points in the region of 400 nm to 800 nm.

References

Nathan C. Fuenffinger (2015), Multivariate Classification Model Transfer for the Discrimination of Textile Fibers by UV-Visible Microspectrophotometry.

Future Research

The performance of our proposed Bayesian classifier depend on the distribution of curves and the statistical depth function. The accuracy of estimation for the distribution of the depth plays a key role on the success of the Bayesian classifier. The ongoing research is applying Monte Carlo Markov Chain method to address the estimation of the distribution of the depth.

Contact Information

Web http://www.people.stat.sc.edu/chongm/
Email chongm@email.sc.edu
Phone (419) 806 9810