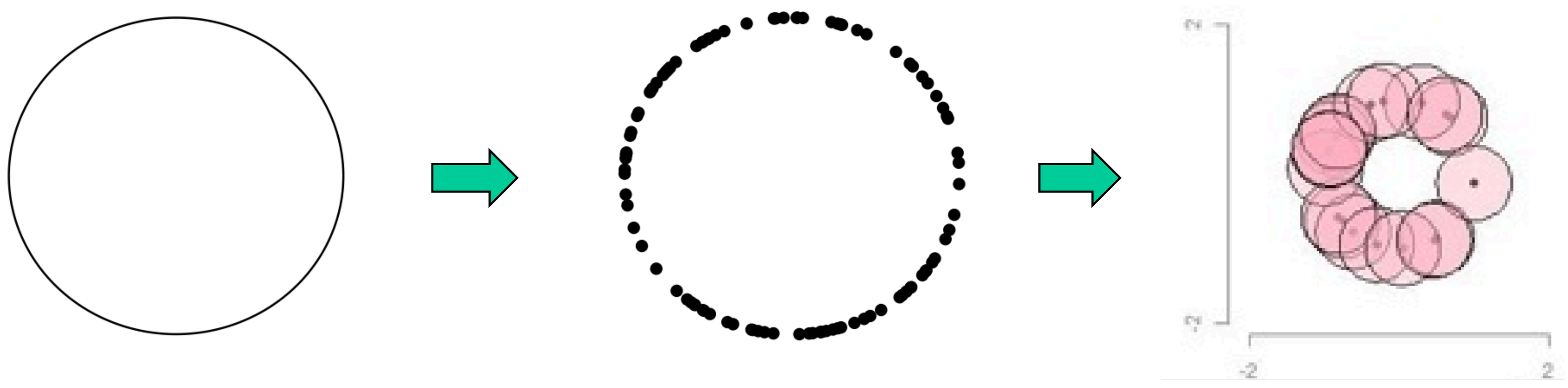
2017年6月16日 統計数理研究所 オープンハウス

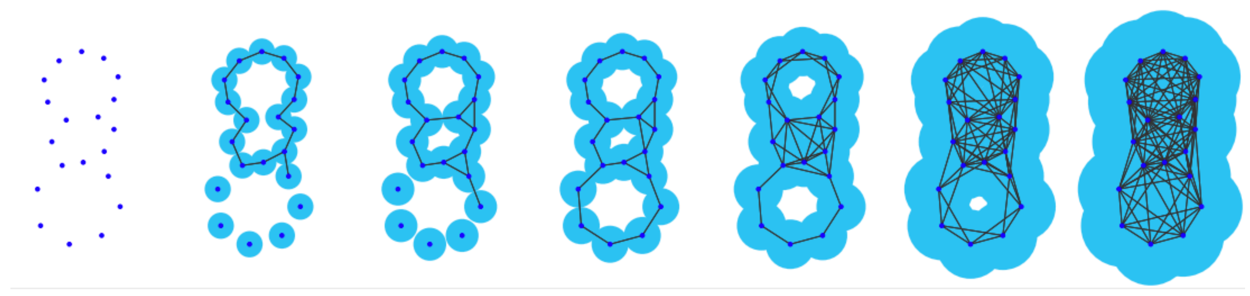
Kernel Methods in Topological Data Analysis

Jin Zhou 統計的機械学習研究センター 特任研究員 （ISM・福水健次先生との共同研究）

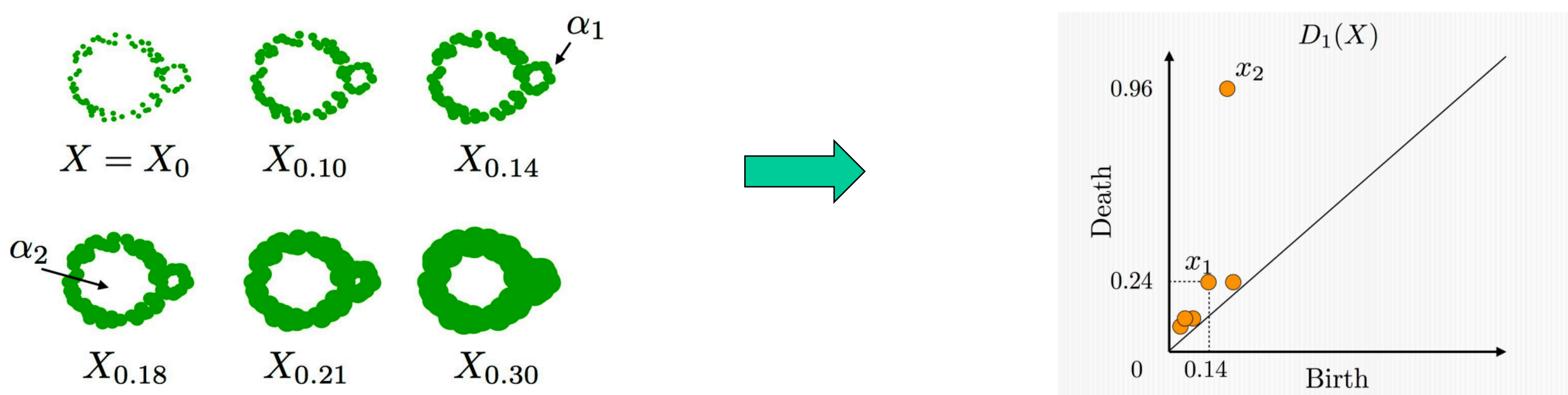
❑ Topological Data Analysis

- Extract the geometrical and topological information of complex data set.
- High dimensional data -> finite sample (point cloud) -> Topological structure





- Persistent Diagram* (Birth-Death time)



❑ Probability Measure on PD and Kernel Embedding

- Statistical inference on PD needs probability measures that support means, expectation, covariance…
- Wasserstein distance

$$W_p(d_1, d_2) = \left(\inf_{\gamma} \sum_{x \in d_1} \|x - \gamma(x)\|_\infty^p \right)^{\frac{1}{p}}$$

- Space of persistence diagrams

$$D_p = \{d \mid W_p(d, d_\emptyset) < \infty\} = \{d \mid \mathrm{Pers}_p(d) < \infty\}$$

- Frechet expectation

$$\mathrm{Var}_{\mathcal{P}} = \inf_{d \in D_p} \left[F_{\mathcal{P}}(d) = \int_{D_p} W_p(d,e)^2 \, \mathrm{d}\mathcal{P}(e) < \infty \right]$$

- Or kernel way, PD can be embedded into RKHS by Bocher integral (vectorization)

$$E_k : M_b(\Omega) \to \mathcal{H}_k, \quad \mu \mapsto \int k(\cdot, x) d\mu(x)$$

- Uniqueness

If k is C_0-universal, the mapping E_k is injective.

$$d_k(\mu, \nu) = \|E_k(\mu) - E_k(\nu)\|_{\mathcal{H}_k}$$

- defines a distance on Mb

❑ Persistence weighted Gaussian kernel

- Distance measure

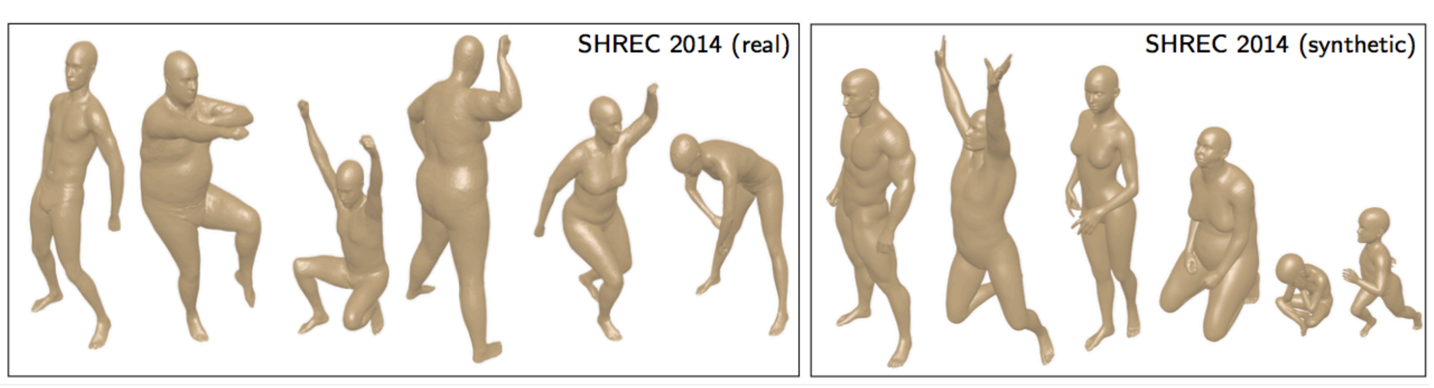
$$K_L(D,E) = \langle E_{k_G}(\mu_D^{w_{\mathrm{arc}}}), E_{k_G}(\mu_E^{w_{\mathrm{arc}}}) \rangle_{\mathcal{H}_{k_G}} = \sum_{x \in \underline{D}} \sum_{y \in \underline{E}} w_{\mathrm{arc}}(x) w_{\mathrm{arc}}(y) k_G(x,y)$$

- For k(,), Gaussian kernel is used.
- Weight is defined by

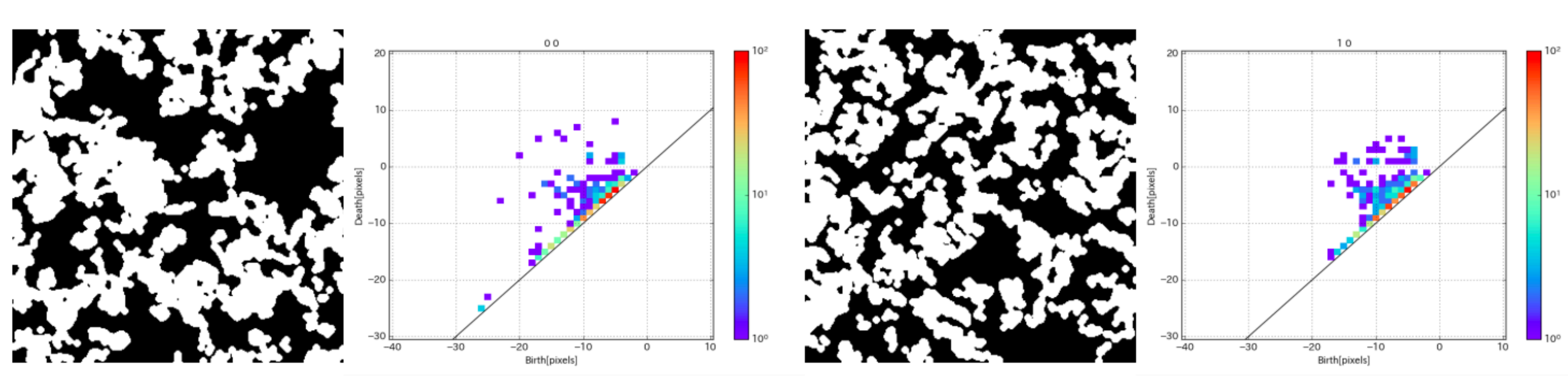
$$w_{\mathrm{arc}}(x) = \arctan(C \mathrm{pers}(x)^p) \ (C, p > 0)$$

❑ Applications

- Computer vision



- Material Science



Reference: 1.Persistence weighted Gaussian kernel for topological data analysis. Genki Kusano, Kenji Fukumizu, Yasuaki Hiraoka
*Images are taken from the above publication.
2.Probability measures on the space of persistence diagrams. Yuriy Mileyko, Sayan Mukherjee and John Harer