An invariant for the comparison of composition operators for reproducing kernel Hilbert spaces

Isao Ishikawa
(j/w Keisuke Fujii, Masahiro Ikeda, Yuka Hashimoto, Yoshinobu Kawahara)

RIKEN Center for Advanced Intelligence Project / Keio University

6/27/2018

Contents

Contents

Motivation

Classifying or clustering time-series data is **important**

Human activities

data1

$t = 0$ $t = 1$ $t = 2$

data2

$t = 0$ $t = 1$ $t = 2$

Nature

data1

$t = 0$ $t = 1$ $t = 2$

data2

$t = 0$ $t = 1$ $t = 2$

Problem

How to compare time series datas ??

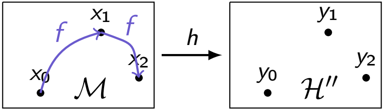
Strategy

Given time series data in a Hilbert space $\mathcal{H}''$: $y_0, y_1, y_2, \ldots, y_t, \cdots \in \mathcal{H}''$

Assume $\{y_t\}_{t \geq 0}$ is generated by a **dynamical system**

$f : \mathcal{M} \to \mathcal{M}$, $\{x_t\} \subset \mathcal{M}$
$h : \mathcal{M} \to \mathcal{H}''$

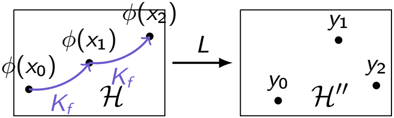
$$x_{t+1} = f(x_t)$$
$$y_t = h(x_t)$$



$\rightsquigarrow$

$\mathcal{H}$: RKHS
$\phi : \mathcal{M} \to \mathcal{H}$: feature map
$K_f : \mathcal{H} \to \mathcal{H}$: Koopman operator
$L_h : \mathcal{H} \to \mathcal{H}''$: bounded linear

$$\phi(x_{t+1}) = K_f \phi(x_t)$$
$$y_t = L_h \phi(x_t)$$



In terms linear algebra ($\mathcal{H}$, K_f, L_h e.t.c), we can define a "good" invariant !

Contents

Positive definite kernel and RKHS

$\mathcal{M}$: set

Definition

$k\colon \mathcal{M} \times \mathcal{M} \longrightarrow \mathbb{C}$: positive definite kernel if

for any $n > 0$ and $x_1, \ldots, x_n \in \mathcal{M}$,

$$\big(k(x_i, x_j)\big)_{i,j=1,\ldots,n}$$

is a positive semi-definite Hermitian matrix.

$k_a\colon \mathcal{M} \to \mathbb{C}$: $k_a(x) := k(x, a)$ $(a \in \mathcal{M})$.

$$\phi\colon \mathcal{M} \to \mathbb{C}^{\mathcal{M}}; x \mapsto k_x\text{: feature map}$$

Theorem (Moore-Aronszajn)

k: positive definite kernel on $\mathcal{M}$

$$\exists! \mathcal{H} \subset \mathbb{C}^{\mathcal{M}}: \text{ \textbf{reproducing kernel Hilbert space (RKHS)}}$$

characterized by

$$\begin{cases} \bullet \quad \mathcal{H} = \overline{\sum_{a \in \mathcal{M}} \mathbb{C} k_a} = \overline{\operatorname{span}\left(\phi(\mathcal{M})\right)} \\ \bullet \quad \langle k_a,\, k_b \rangle_{\mathcal{H}} = k(a, b) \end{cases}$$

Examples of RKHS

Example

1. $\mathcal{M} = \mathbb{R}^d$
$$k(x, y) := x \cdot y$$
$$\mathcal{H} := \mathrm{Hom}_{\mathbb{R}}(\mathbb{R}^d, \mathbb{C}) \cong \mathbb{C}^N$$

2. (Gaussian kernel) $\mathcal{M} = \mathbb{R}^d$
$$k(x, y) := e^{-|x-y|^2/\sigma^2}$$
$$\mathcal{H} := \left\{ f \in L^2(\mathbb{R}^d) \;\middle|\; \int_{\mathbb{R}^d} \left|\widehat{f}(x)\right|^2 e^{\pi^2\sigma^2|x|^2}\, dx < \infty \right\}$$

3. (Szegö kernel) $\mathcal{M} = \mathbb{D} := \{z \in \mathbb{C} \mid |z| < 1\}$,
$$k(z, w) := (1 - z\overline{w})^{-1}$$
$$\mathcal{H} := \left\{ f\colon \text{regular on } \mathbb{D} \;\middle|\; \sup_{0<r<1} \int_{|z|=r} |f(z)|^2\, dz < \infty \right\}$$

Contents

Definition of Koopman operators for feature maps

k: positive definite kernel on $\mathcal{M}$

$\mathcal{H}$: RKHS

$\mathcal{N} \subset \mathcal{M}$: subset

$$\mathcal{H}_{\mathcal{N}} := \overline{\mathrm{span}(\phi(\mathcal{N}))}$$

Definition (Kawahara)

$f : \mathcal{N} \to \mathcal{N}$: map

The Koopman operator for the feature map is a linear operator

$$K_f : \mathcal{H}_{\mathcal{N}} \longrightarrow \mathcal{H}_{\mathcal{N}}$$

satisfying

1. K_f: bounded
2. For any $y \in \mathcal{N}$, $K_f(\phi(y)) = \phi(f(y))$

If K_f exists, K_f is uniquely determined.

Examples of Koopman operators for feature maps

Example

1. $\mathcal{M} = \mathbb{R}^d$, $\mathcal{N} \subset \mathbb{R}^d$: open, $k(x, y) = e^{-|x-y|^2/\sigma^2}$
   For $f : \mathcal{N} \to \mathcal{N}$: linear map,

   $$K_f \text{ exists} \iff f\text{: invertible \& } ||f|| \leq 1$$

2. $\mathcal{M} = \mathcal{N} = \mathbb{D}$, $k(z, w) = (1 - z\bar{w})^{-1}$
   For $f : \mathbb{D} \to \mathbb{D}$: holomorphic map, by Littlewood,

   $\rightsquigarrow K_f$ exists

3. (Fock space)$\mathcal{M} = \mathcal{N} = \mathbb{C}^d$, $k(z, w) = e^{z \cdot w/2}$
   For $f : \mathbb{C}^d \to \mathbb{C}^d$: holomorphic map, by Carswell-MacCluer-Schuster,
   K_f exists

   $$\iff f(z) = Az + b\text{: affine map} \quad \text{s.t.} \begin{cases} ||A|| \leq 1 \\ A\zeta \cdot b = 0 \text{ for } |A\zeta| = |\zeta| \end{cases}$$

Problem

Whether K_f exists or not is difficult !!

Remark

When $\mathcal{N} = \mathcal{M}$,

$$K_f \text{ exists} \iff \text{composition operator for } f \text{ is bounded}$$

With example 1, there are no example of a non-linear dynamical system such that its Koopman operator exists ! (in progress)

Contents

Generalized observable

$\mathcal{H}$: RKHS with feature map $\phi : \mathcal{M} \to \mathcal{H}$
$\mathcal{N} \subset \mathcal{M}$: subset

Definition

$\mathcal{H}''$: Hilbert space
$h : \mathcal{N} \to \mathcal{H}''$: generalized observable on $\mathcal{H}_{\mathcal{N}}$ if
$\exists L_h : \mathcal{H}_{\mathcal{N}} \to \mathcal{H}''$: bounded linear s.t.

$$h = L_h \circ \phi|_{\mathcal{N}}$$

Examples of generalized observable

1. $h = \phi|_{\mathcal{N}} : \mathcal{N} \to \mathcal{H}$

$$L_h : \mathcal{H}_{\mathcal{N}} \hookrightarrow \mathcal{H}: \text{ inclusion}$$

2. $h(x) = (g_i(x))_{i=1}^N : \mathcal{N} \to \mathbb{C}^N \ (g_i \in \mathcal{H})$

$$L_h(v) = (\langle g_i, v \rangle_{\mathcal{H}})_{i=1}^N$$

Setting

$\mathcal{M}_1, \mathcal{M}_2 \subset \mathcal{M}$: subsets
$\mathcal{H}', \mathcal{H}''$: Hilbert spaces

For $i = 1, 2$,

$$
\begin{aligned}
f_i : &\quad \mathcal{M}_i \longrightarrow \mathcal{M}_i &&: \text{map (dynamical system)} \\
h_i : &\quad \mathcal{M}_i \longrightarrow \mathcal{H}'' &&: \text{generalized observable} \\
X_i : &\quad \mathcal{H}' \longrightarrow \mathcal{H}_{\mathcal{M}_i} &&: \text{Hilbert-Schmidt operator (initial value)}
\end{aligned}
$$

Assume for $i = 1, 2$,

$$
\exists K_{f_i} : \mathcal{H}_{\mathcal{M}_i} \longrightarrow \mathcal{H}_{\mathcal{M}_i}
$$
$$
\exists L_{h_i} : \mathcal{H}_{\mathcal{M}_i} \longrightarrow \mathcal{H}''
$$

Definition of the "angle" between Koopman operators

For $T, m > 0$,

$$d_m^T\left((L_{h_1}, K_{f_1}, X_1), (L_{h_2}, K_{f_2}, X_2)\right) := \mathrm{tr}\left(\bigwedge^m \sum_{r=1}^{T} \left(L_{h_2} K_{f_2}^r X_2\right)^* L_{h_1} K_{f_1}^r X_1\right)$$

$$A_m^T\left((h_1, f_1, X_1), (h_2, f_2, X_2)\right) := \frac{\left|d_m^T\left((L_{h_1}, K_{f_1}, X_1), (L_{h_2}, K_{f_2}, X_2)\right)\right|^2}{\prod_{i=1}^{2} d_m^T\left((L_{h_i}, K_{f_i}, X_i), (L_{h_i}, K_{f_i}, X_i)\right)}$$

$$\widetilde{A}_m^T := \frac{1}{T} \sum_{t=1}^{T} A_m^t$$

Definition (The angle of dynamical systems)

If $\lim_T \widetilde{A}_m^T((h_1, f_1, X_1), (h_2, f_2, X_2))$ exists, we define

$$A_m((h_1, f_1, X_1), (h_2, f_2, X_2)) := \lim_{T\to\infty} \widetilde{A}_m^T((h_1, f_1, X_1), (h_2, f_2, X_2))$$

Remark

We prove some sufficient conditions for the convergence of $\widetilde{A}_m^T$ in the case of $\mathcal{H}' = \mathbb{C}^N$.

Contents

Comparison of state space models

$\mathcal{M} = \mathcal{M}_1 = \mathcal{M}_2 = \mathbb{R}^N$, $k(x, y) = x \cdot y$, $\mathcal{H}' = \mathcal{H}$, $\mathcal{H}'' = \mathbb{C}^M$

For $i = 1, 2$, consider a state space model (without noise):

$$\mathbf{x}_{t+1} = \mathbf{A}_i \mathbf{x}_t$$
$$\mathbf{y}_t = \mathbf{C}_i \mathbf{x}_t$$

where $\mathbf{A}_i \in \mathbb{R}^{N \times N}$, $\mathbf{C}_i \in \mathbb{R}^{M \times N}$

Theorem

For any $m > 0$,

$$\lim_{T\to\infty} \widetilde{A}_m^T((\mathbf{C}_1, \mathbf{A}_1, \mathbf{I}_N), (\mathbf{C}_2, \mathbf{A}_2, \mathbf{I}_N)) \text{ exists}$$

Remark

In the case $m = N$ and state space models are stable (i.e. the eigenvalues of $\mathbf{A}_i$ are smaller than 1), DeCock-DeMoor define an angle between the state space models, which is the same as $A_N((\mathbf{C}_1, \mathbf{A}_1, \mathbf{I}_N), (\mathbf{C}_2, \mathbf{A}_2, \mathbf{I}_N))$.

Comparison of AR models

$\mathcal{M} = \mathcal{M}_1 = \mathcal{M}_2 = \mathbb{R}^N$, $k(x, y) = x \cdot y$, $\mathcal{H}' = \mathcal{H}$, $\mathcal{H}'' = \mathbb{C}$

Two AR models of order N ($i = 1, 2$):

$$(M_i) \quad x_t = a_{i,1}x_{t-1} + \cdots + a_{i,N}x_{t-N}$$

$$\Updownarrow$$

Two state space models:

$$\mathbf{x}_t = \mathbf{A}_i \mathbf{x}_{t-1},$$
$$y_t = \mathbf{C}_i \mathbf{x}_t$$

$$\mathbf{A}_i = \begin{pmatrix} a_{i,1} & \cdots & a_{i,N} \\ \mathbf{I}_{N-1} & & \mathbf{o} \end{pmatrix}, \quad \mathbf{C}_i = (1, 0, \ldots, 0)$$

Let's compute $A_N((\mathbf{C}_1, \mathbf{A}_1, \mathbf{I}_N), (\mathbf{C}_2, \mathbf{A}_2, \mathbf{I}_N))$!

$\gamma_{i,1}, \ldots, \gamma_{i,N}$: roots of $X^N - a_{i,1}X^{N-1} - \cdots - a_{i,N} = 0$

$$P_i := \left\{\gamma_{i,n} \;\middle|\; |\gamma_{i,n}| > 1\right\}, \; Q_i := \left\{\gamma_{i,n} \;\middle|\; |\gamma_{i,n}| = 1\right\}$$
$$R_i := \left\{\gamma_{i,n} \;\middle|\; |\gamma_{i,n}| < 1\right\}.$$

Theorem

If $|P_1| = |P_2|$, $|R_1| = |R_2|$, and $Q_1 = Q_2$, we have

$$A_N\left((\mathbf{C}_1, \mathbf{A}_1, \mathbf{I}_N), (\mathbf{C}_2.\mathbf{A}_2, \mathbf{I}_N)\right)$$
$$= \frac{\prod_{\alpha,\beta \in P_1} (1 - \alpha\overline{\beta}) \cdot \prod_{\alpha,\beta \in P_2} (1 - \alpha\overline{\beta})}{\prod_{\alpha \in P_1, \beta \in P_2} |1 - \alpha\beta|^2} \cdot \frac{\prod_{\alpha,\beta \in R_1} (1 - \alpha\overline{\beta}) \cdot \prod_{\alpha,\beta \in R_2} (1 - \alpha\overline{\beta})}{\prod_{\alpha \in R_1, \beta \in R_2} |1 - \alpha\beta|^2},$$

otherwise, $A_N((\mathbf{C}_1, \mathbf{A}_1, \mathbf{I}_N), (\mathbf{C}_2, \mathbf{A}_2, \mathbf{I}_N)) = 0$

Remark

If M_i's are stable (i.e. $P_i = Q_i = \emptyset$), it is the same as the metric between M_i's defined by Martin via cepstrum coefficients.

A formula for computations

$\mathcal{H}' = \mathbb{C}^N$, $\mathcal{H}'' = \mathcal{H}$
$h_i = \phi|_{\mathcal{M}_i}$

For $i = 1, 2$, N time series sequences are given:

$$\begin{gathered}\{x^1_{i,0}, x^1_{i,1}, x^1_{i,2}, \dots\} \subset \mathcal{M}_i \\ \{x^2_{i,0}, x^2_{i,1}, x^2_{i,2}, \dots\} \subset \mathcal{M}_i \\ \vdots \\ \{x^N_{i,0}, x^N_{i,1}, x^N_{i,2}, \dots\} \subset \mathcal{M}_i\end{gathered}$$

Assume

$$f_i\left(x^r_{i,t}\right) = x^r_{i,t+1}.$$

$X_i : \mathbb{C}^N \to \mathcal{H}$

$$X_i((0, \dots, 0, \overset{r}{1}, 0, \dots, 0)) := \phi(x^r_{i,0}).$$

For $m = 1, \ldots, N$,

Theorem

$$d_m^T((L_{h_1}, K_{f_1}, X_1), (L_{h_2}, K_{f_2}, X_2))$$
$$= \sum_{t_1,\ldots t_m=0}^{T} \sum_{0<s_1<\cdots<s_m\leq N} \det \begin{pmatrix} k\left(x_{i,t_1}^{s_1}, x_{j,t_1}^{s_1}\right) & \cdots & k\left(x_{i,t_1}^{s_1}, x_{j,t_m}^{s_m}\right) \\ \vdots & \ddots & \vdots \\ k\left(x_{i,t_m}^{s_m}, x_{j,t_1}^{s_1}\right) & \cdots & k\left(x_{i,t_m}^{s_m}, x_{j,t_m}^{s_m}\right) \end{pmatrix}$$

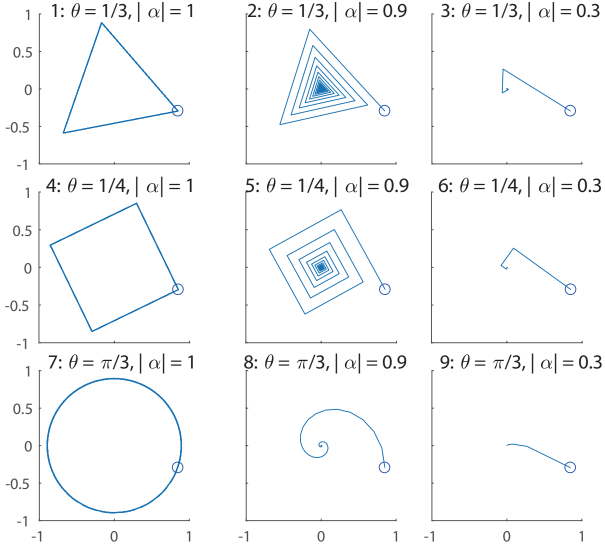
Point

We can compute A_m from only the value of k (kernel trick).

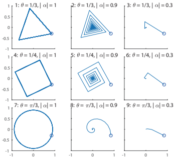
Contents

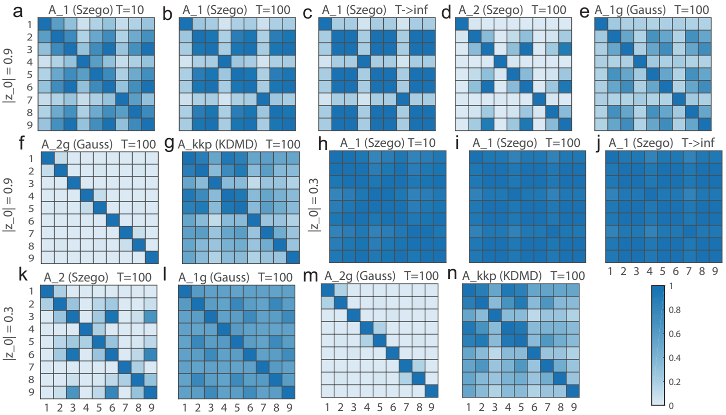
Rotation on the unit disk

Consider 9 dynamical systems (the multiplication of $\alpha = |\alpha|e^{i\theta}$ on $\mathbb{D}$):



Computational result





m	data
1	$z_0, \alpha z_0, \ldots$
2	$z_0, \alpha z_0, \ldots$ $\alpha z_0, \alpha^2 z_0, \ldots$

A_m (Szego) T= ? = $A_m^?$ w.r.t Szegö

A_1 (Szego) T->inf = $\lim_T A_1^T$ w.r.t Szegö

A_mg (Gauss) T= ? = $A_m^?$ w.r.t Gaussian

A_kkp (KDMD) = other method

Thank you for listening !

For detail, check arXiv 1805.12324 !