

# カーネル平均を用いた記述子設計

青木 祐太

ものづくりデータ科学研究センター 特任助教

## 背景 集合データに対するデータ分析

- ✓ 要素数がサンプルによって異なる。

$$S_i = \{s_i^{(1)}, s_i^{(2)}, \dots, s_i^{(p)}\}$$

$$S_j = \{s_j^{(1)}, s_j^{(2)}, \dots, s_j^{(q)}\}$$

- ✓ 任意の要素間の交換に対する不変性。  
(順不同)

$$S_i = \{s_i^{(1)}, s_i^{(2)}, \dots, s_i^{(k)}, \dots, s_i^{(l)}, \dots, s_i^{(p)}\}$$

$$= \{s_i^{(1)}, s_i^{(2)}, \dots, s_i^{(l)}, \dots, s_i^{(k)}, \dots, s_i^{(p)}\}$$

通常の機械学習の枠組みに載せるには、集合データ  $S_i$  を固定長のベクトル記述子  $x_i$  に変換する必要がある。 → カーネル平均の利用

## 手法 カーネル平均

$$m_X(\cdot) := E_{X \sim P}[k(\cdot, X)]$$

- ✓ カーネル平均の性質:
  - $X$ の分布 $P$ に関する高次モーメントの情報を持つ。
  - カーネル $k$ が「特性的」であるとき、 $P$ を一意に定める。



## カーネル平均記述子

$$x_i(\cdot) := \hat{m}_{S_i}(\cdot) = \frac{1}{p} \sum_{n=1}^p k(\cdot, s_i^{(n)})$$

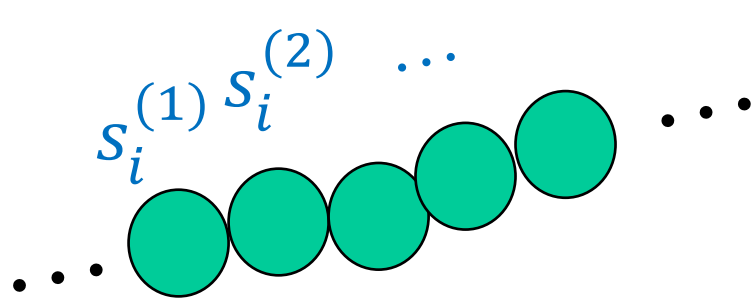
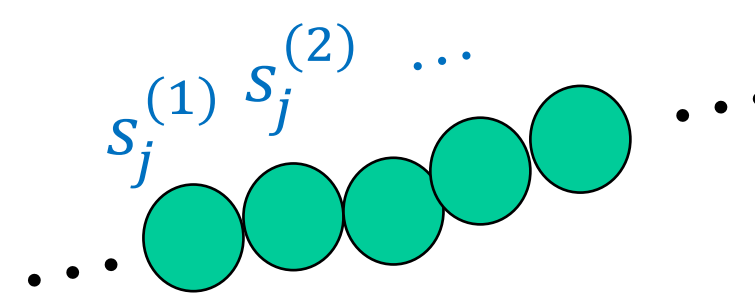
$x_i(\cdot)$ は関数なので、離散化して固定長ベクトルとする。

ハイパーパラメータ: 離散化の範囲, メッシュ間隔, カーネルのパラメータ

採用したカーネル: ガウスカーネル  $k(x, y) = \exp\left[-\frac{\|x - y\|^2}{2\sigma}\right]$

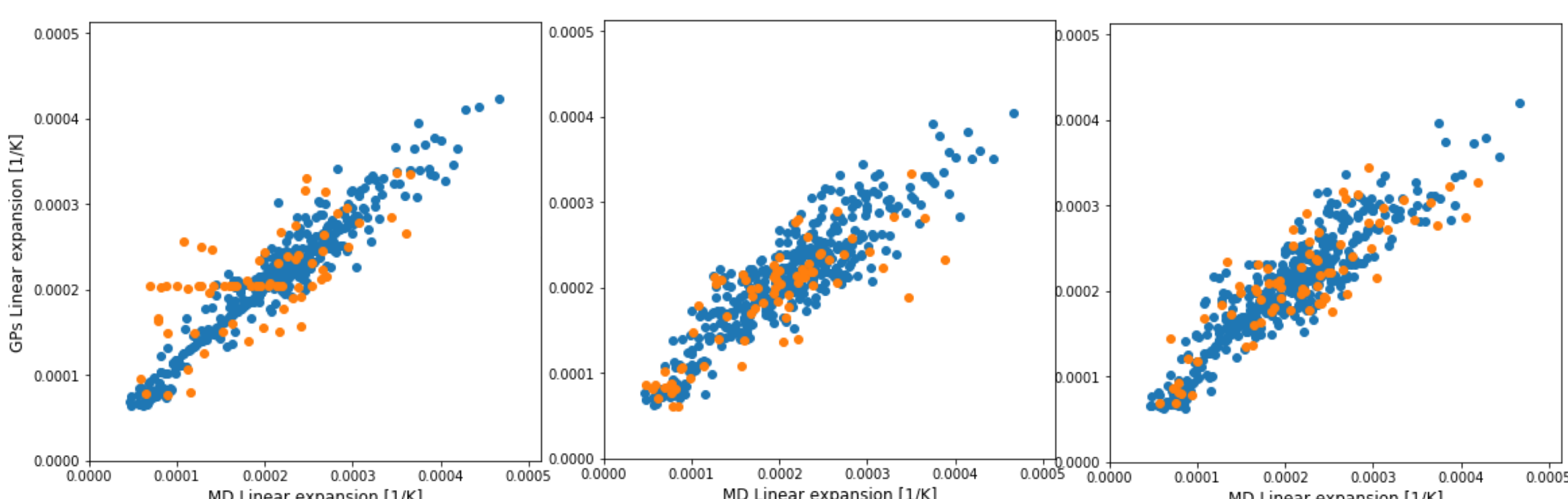
## 応用例1 高分子物性の予測

(林慶浩 特任助教)

高分子 $i$ 高分子 $j$ 

$s_i^{(n)}$  ... 高分子 $i$ を構成する原子 $n$ を表現するパラメータ  
分子力学のカ場パラメータ9種類<sup>※3</sup>

### 熱膨張係数に対する予測結果 (ガウス過程回帰)

RDKit記述子<sup>※1</sup>Sに対する要約統計量<sup>※2</sup>Sに対するカーネル平均<sup>※3</sup>

※1 RDKit記述子はchemoinformatics分野で使われる代表的な分子記述子の一つ。

<https://rdkit.org/>

※2 要約統計量として平均, 標準偏差, 歪度, 尖度, 最大値, 最小値の6種類を採用。

※3 カ場パラメータ9種類に対して個別にカーネル平均記述子を作成し, それら全てを使用。

## 応用例2 準結晶となる化学組成の予測

(劉暢 特任助教)

組成 $i$  $s_i^{(1)} s_i^{(2)} s_i^{(3)}$  $Zn_\alpha Mg_\beta Sc_\gamma$ 組成 $j$  $s_j^{(1)} s_j^{(2)} s_j^{(3)} s_j^{(4)}$  $Cu_{\alpha'} Ga_{\beta'} Mg_{\gamma'} Sc_{\delta'}$ 

$s_i^{(n)}$  ... 組成 $i$ を構成する元素 $n$ を表現するパラメータ

元素に対して定義, 測定, 算出できる元素物性値58種類



組成比 $\alpha, \beta, \gamma, \dots$ で加重平均をとってカーネル平均記述子を算出

### 準結晶/近似結晶/その他(結晶等)の分類予測精度 (ランダムフォレスト)

XenonPy組成記述子<sup>※4</sup> (要約統計量: 加重平均, 加重分散, 最大値, 最小値)

クラス	Recall	Precision	F <sub>1</sub> score
準結晶	0.602	0.722	0.650
近似結晶	0.608	0.731	0.658
その他(結晶等)	0.999	0.997	0.998

カーネル平均記述子 (58種類の元素物性値に対して個別にカーネル平均記述子を作成し, それら全てを使用)

準結晶	0.657	0.766	0.703
近似結晶	0.637	0.744	0.681
その他(結晶等)	0.999	0.997	0.998

※4 <https://xenonpy.readthedocs.io/en/latest/>