

土木学会 AI・データサイエンスセミナー

テキストとその背後の ダイナミクスをとらえる

持橋大地

統計数理研究所 数理・推論研究系

daichi@ism.ac.jp

2023-10-19 (木)
オンライン講演

自己紹介

立川・統数研

- 持橋大地
統計数理研究所 数理・推論
研究系
 - 1998年 東京大学教養学部
基礎科学科第二卒業
 - 2005年 奈良先端科学技術大学院大学 情報科学
研究科 博士後期課程修了
- 専門：自然言語処理、統計的機械学習
言語に限らず、多数の分野と共同研究を行っています



土木工学との関わり

- 大学の同級生(文科)が土木に定数外で進学
- 個人的に、建築・土木にも興味あり
 - 小林潔司先生の階層的隠れマルコフ舗装劣化モデル(2011)について知ったのがきっかけ

↓ 東京スカイツリーに初めて行った後に読んだ論文

TS29-04

Structural Design of Tokyo Sky Tree

Atsuo konishi¹

Senior Structural Engineer, Structural Engineering Department, Nikken Sekkei Ltd., Tokyo, Japan
konishi@nikken.co.jp



Atsuo Konishi

Atsuo KONISHI, born in 1963, received his Master of Architectural Engineering degree from the University of Kobe, Japan. He has 20 years experience in building and tower design as a structural engineer, and is currently a senior structural engineer for Nikken Sekkei Ltd. He was a structural designer of Tokyo sky tree, and the structural design was begun in 2004.

Abstract

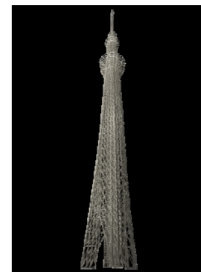
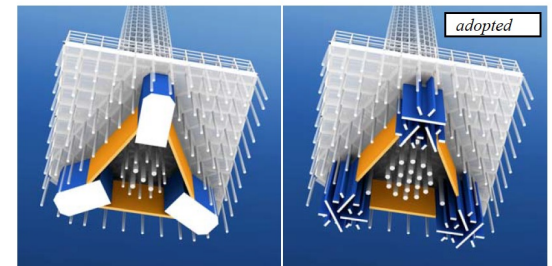


Figure 2 Superstructure



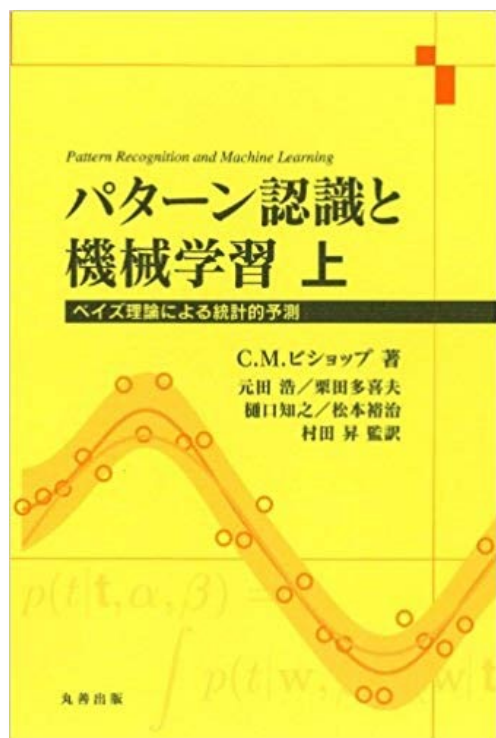
Counterweight plan SRC(bl.)and RC(ye.) wall pile plan
Figure 3 : Plan for Substructure

The wall pile plan (Fig.3: right plan, adopted) uses the weight of soil grasped by foundation as pull-out resistance, using the friction of the pile surface. The counterweight plan is simple and reliable system compared with the wall pile plan. But wall pile plan was adopted in the basic design, because it excels at the adjacent construction near the subway.

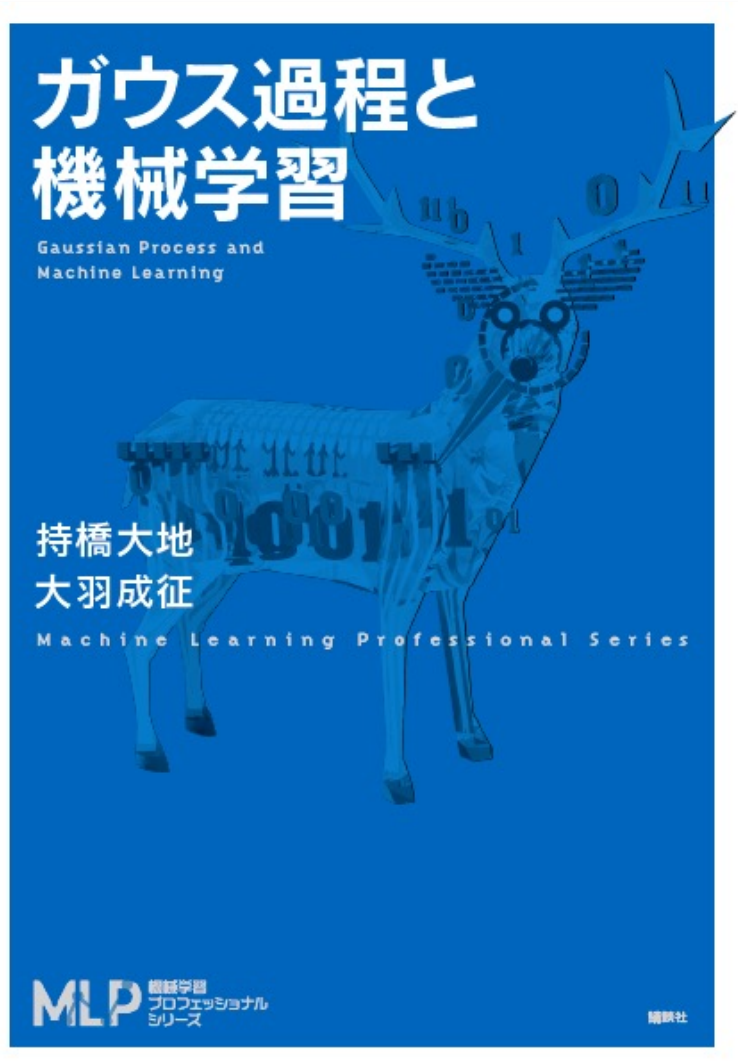
The aspect ratio of Tokyo Sky Tree is about 9.0, and the top displacement is too large to broadcast if the foundation has poor vertical rigidity because of the uplift-rocking deflection of Foundation. For example, 10.0cm displacements of a pile convert 90.0cm displacements at the top of the tower. A SRC pile (steel reinforced concrete pile, concrete encased steel pile) was adopted as the foundation structure to consider the continuity of steel member as the steel tower. However, behaviours of SRC pile under the pull-out force

関わった教科書

- 『パターン認識と機械学習』 翻訳チーム (2006年頃～)
 - 10章(下巻)の変分ベイズ法の翻訳を担当
 - 今でも、機械学習の最も標準的な教科書



『ガウス過程と機械学習』



- 講談社機械学習プロフェッショナルシリーズ(MLP), 2019年発売
- 現在、Amazonのレビュー数 **122個**
- 線形回帰モデルの非常にやさしい導入から入っています

岩波データサイエンス 自然言語処理



- 「岩波データサイエンス Vol.2 統計的自然言語理」
- 私が特集総括と記事の執筆をしています
- 1500円で読みやすい読み物形式です
- 大きな書店には置いてあると思います

土木領域での自然言語データ

- 以下は、外部からの想像です [補っていただければ幸いです]
- (膨大な)保守記録
 - 橋梁や道路などの状態を自由記述で書いたデータ
 - 形容詞や副詞などを用いて状態を記述
- 住民からの反応や陳情などのテキストデータ
 - 何が問題なのか、焦点がどう変わってきているのかを人手で客観的に判断するのは難しい
- などなど、**テキストに貴重な情報**が埋まっており、それを統計的に有効に取り出したい

注: 大規模言語モデル(LLM)について

- ChatGPT等のLLMが流行っているが、アイデア出しなどにはよいが、動作がまだまだ不安定
- 工学的に安心・安全に運用するのは**難しい**
 - 内部の動作が**ほとんど解明されていない**
 - **品質保証**が全くなされていない



- LLMに頼らずとも(あるいは、ブラックボックスとして有効に利用して)、信頼できる言語処理を**数理的に実現**したい

今日の話

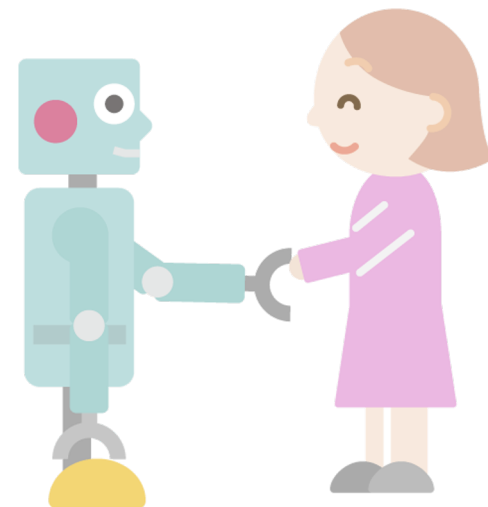
- 周波数空間における「副詞」の数学的モデル化とロボティクスへの応用
- 言語のような離散データの背後に隠れた微分方程式の推定

周波数空間における「副詞」の モデル化

(お茶の水女子大学D2 谷口巴さんとの共同研究)

研究背景

- 生活を支援する家庭用ロボットの重要性の高まり
- 日常生活において、**人と同じ感覚**を共有した動作が期待される
- **副詞**などで表現されるあいまいな言葉を理解し、動作へ結びつけることが必要
 - 「この皿は**丁寧に**運んで」
 - 「庭を**さっと**掃いてから**すぐに**表に回ってね」



使用するデータセット

- 100Walks データ @ YouTube [<https://youtu.be/HEoUhlesN9E>]
 - 「歩く」動作の動画
 - 100個の動画データ、演者1人
 - 100種類の歩く動作
 - 副詞をクラウドでタグ付け
- AIST++ データ
 - 様々な「踊る」動作の動画
 - 1199個の動画データ, 10種類
 - 20人のダンサー
 - 副詞をクラウドでタグ付け



100 Walks 動画データ



AIST++動画データ

副詞をアノテーション

- クラウドソーシングサイトLancersを用いて各動画に副詞のアノテーションを依頼
 - 100 Walks データ：100個ずつ24名に依頼

動画数	副詞の種類	平均副詞数
100	264	12.9

- AIST++ データ：50個ずつ120名に依頼

動画数	副詞の種類	平均副詞数
1199	1767	16.2

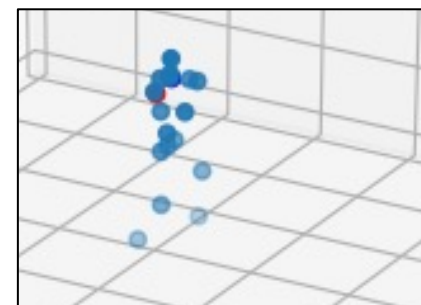
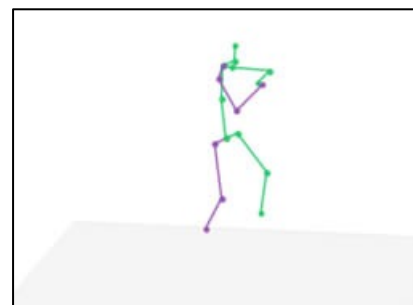
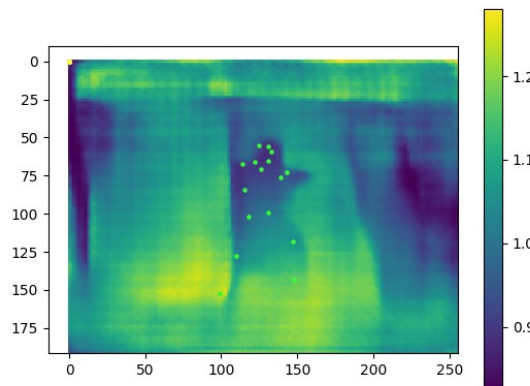
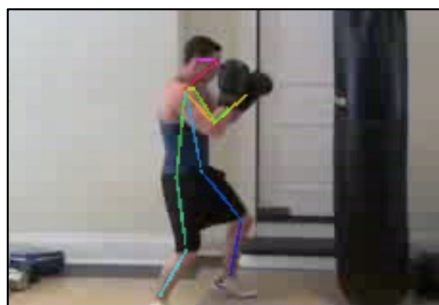
動画にアノテーションされた副詞

	A	B	C	D	E	F	G	H	I	J	K	L
1	lancersTas	lancersRes	lancersNic	taskFormT	taskFormT	taskFormT	taskFormT	taskFormT	taskFormT	taskFormT	taskFormT	taskFormT
				副詞の定義	普通に リラック スして ゆったり と 気だるく	スピー ディーに 普通に 素早く 細やかに さっさと キビキビ と 女性のよ うに 足早に	抜き足差 し足で こっそり と スロー モーショ ンで ゆっくり と 大股で	イライラ と 風を切り ながら さっさと 怒りなが ら かっかと して	ガニ股で 横歩きで 蟹歩きで ちょこま かと かばいな がら ちょこ ちょこと	自信を 持って 足早に 颯爽と 爽快に キビキビ と しなやか に	抜き足差 し足で バレード ンサーの ように 優しく 流れるよ うに	速足で 変わった リズムで テンポ良 く リズムカ ルに 急いで 考えなが ら 素早く
2	LTI004132	LRI030460	Weetabix									
	LTI004132	LRI030460	kei_0107	よろしくお	自然に 気怠そう に	自信あり げに ぐにやぐ にやと	おそろお そろ ゆっくり	力強く 自信あり げに	不自然に 動物的に	自信あり げに 女性っぽ く	ゆっくり おそろお そろ	元気に すばやく
3												

動画データの前処理 (100Walks)

- 動画から骨格の三次元座標を抽出する

もとの動画→



2次元座標を抽出
Open Pose
[Cao+,2018]

深度を推定
FCRN-
DepthPrediction
[Laina+,2016]

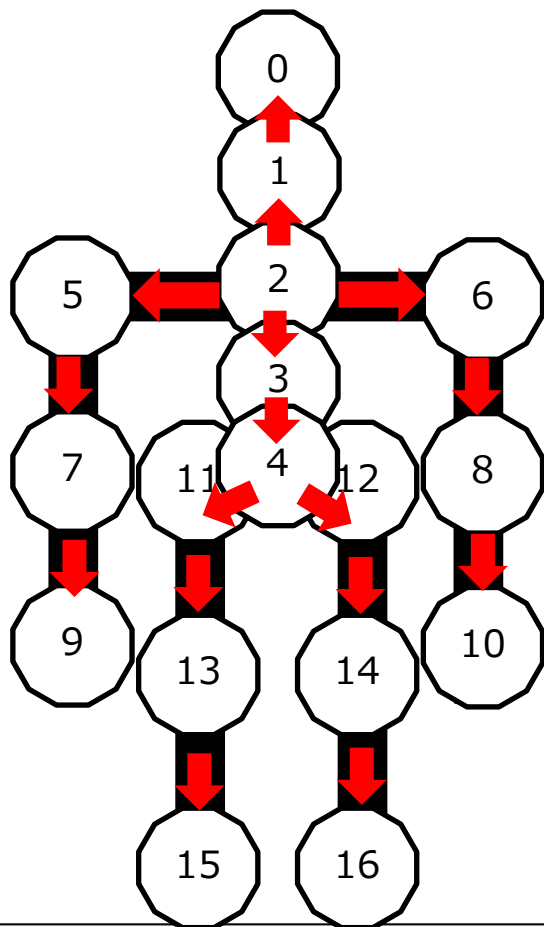
3次元座標を抽出
3d-pose-baseline
[Martnez+,2017]

姿勢を正規化

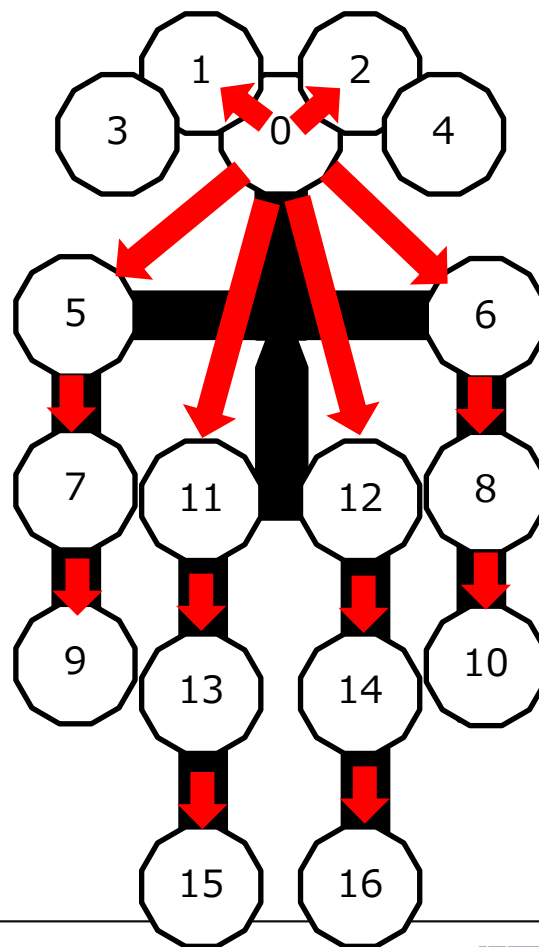
方向ベクトルの算出

- 座標から方向ベクトルを算出

100 Walks データ(16本)



AIST++データ(14本)



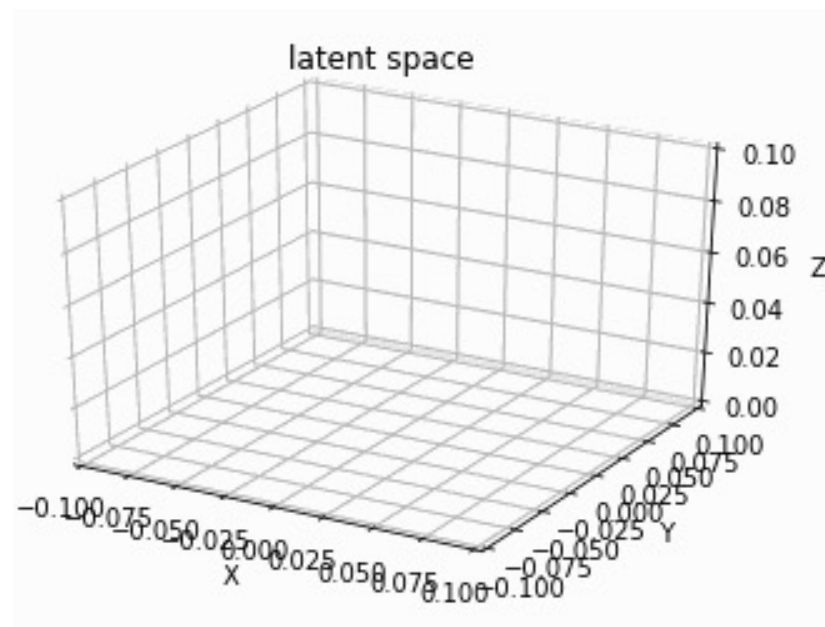
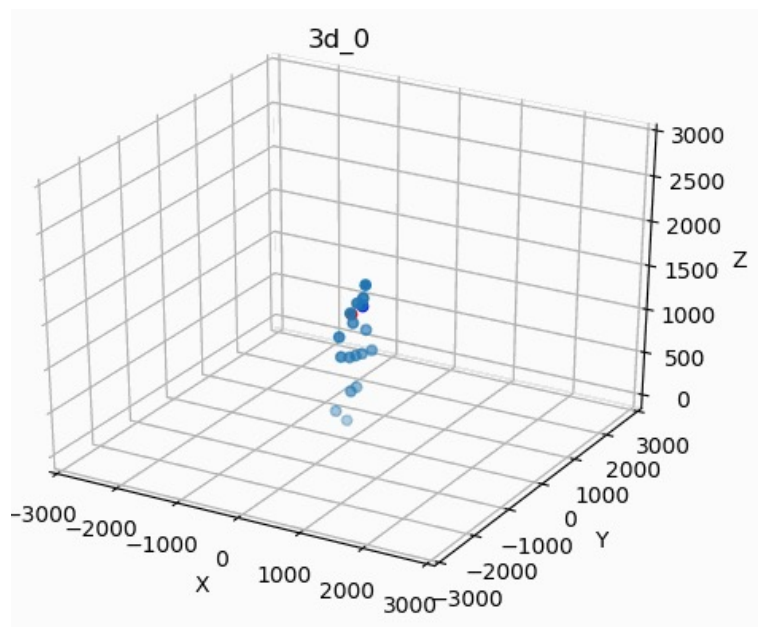
問題1：

- 運動データの次元が高すぎる
 - 100 Walksデータの場合、 $16 \times 3 = 48$ 次元
 - AIST++データの場合、 $14 \times 3 = 42$ 次元
- 次元の間に、高い相関がある
- 真の運動の自由度は、もっと圧倒的に低い!
- 主成分分析 (PCA) をすればよい? → ✕

GPLVMを用いてデータを次元圧縮

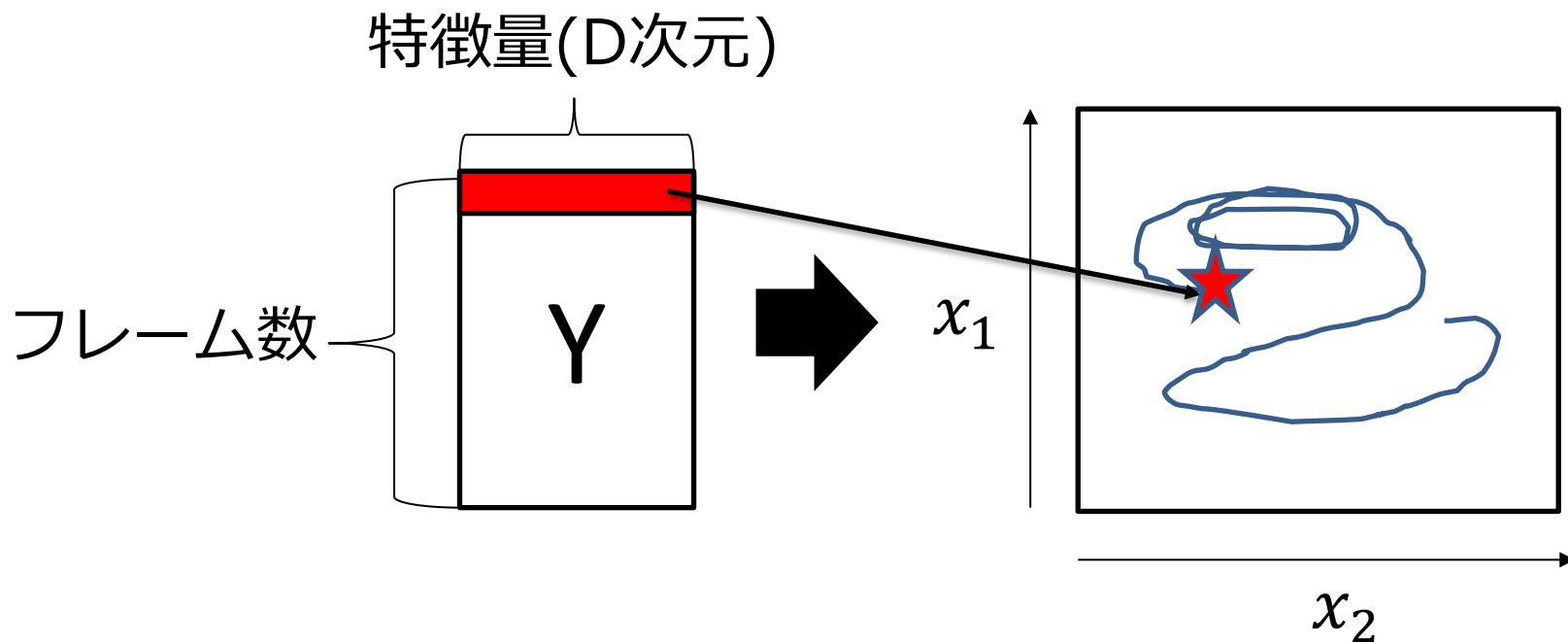
- 48次元の情報を3次元に**非線形**に次元圧縮
- 入力 : $N \times 48$ 行列の姿勢情報
- 出力 : $N \times 3$ 行列の潜在変数

N: フレーム数



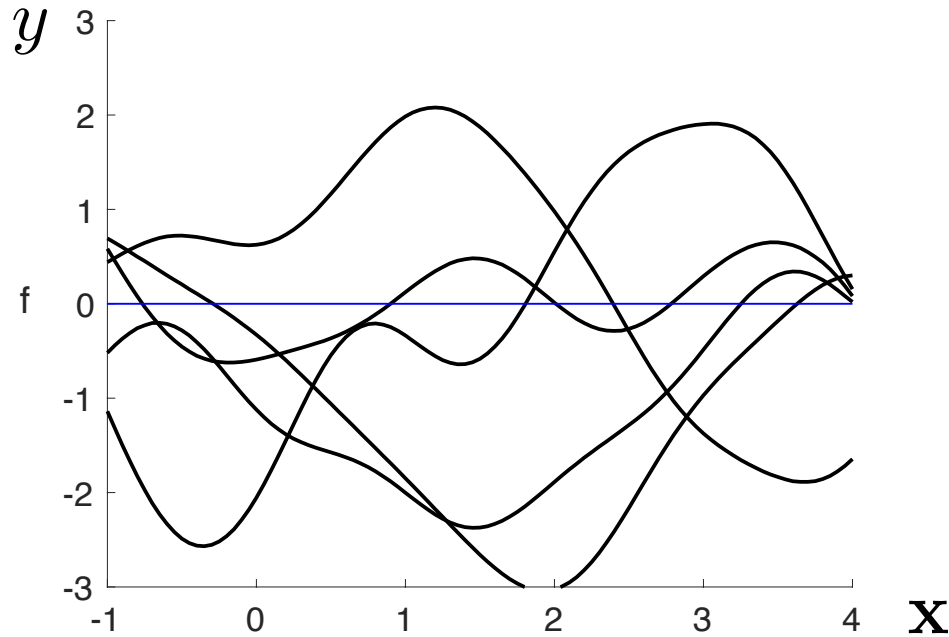
Gaussian Process Latent Variable Model

- ガウス過程を用いた教師なし学習 (Lawrence+ 2003)
- 観測値 \mathbf{Y} が、潜在変数 \mathbf{X} から生成されていると仮定
- カーネル関数を用いて
 \mathbf{X} から \mathbf{Y} への写像を非線形にすることが可能



ガウス過程とは

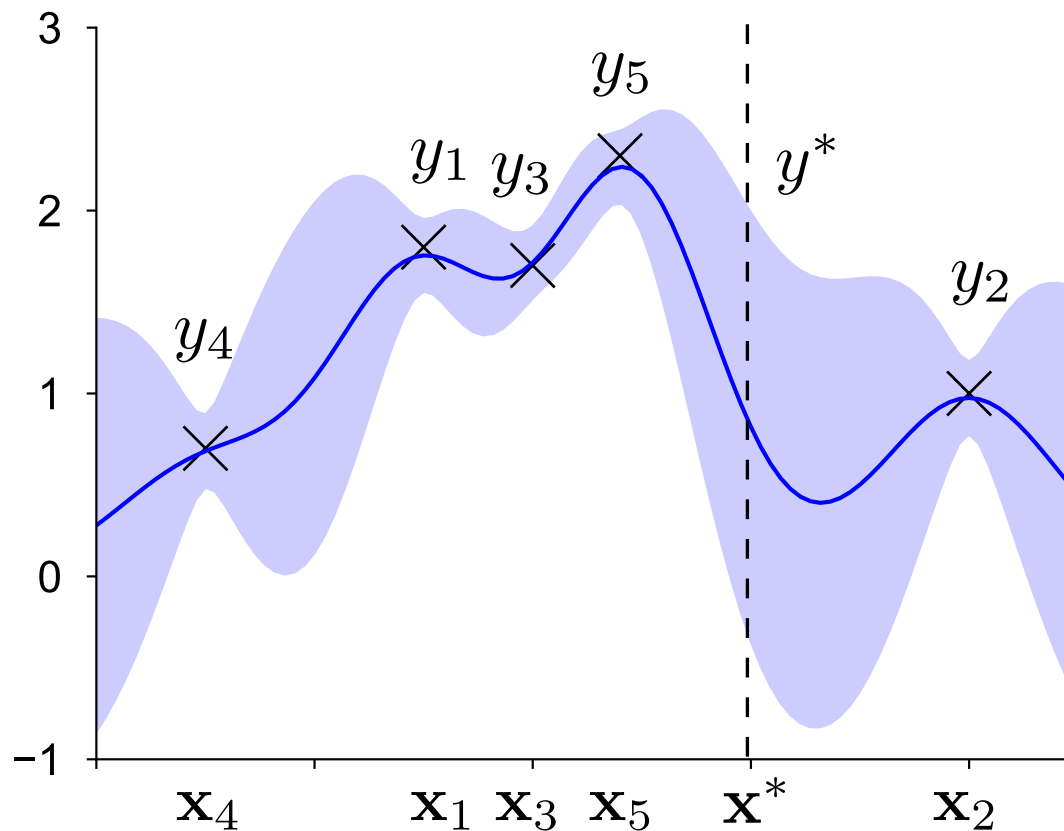
- ガウス過程 (Gaussian process) … 関数を生成する確率分布



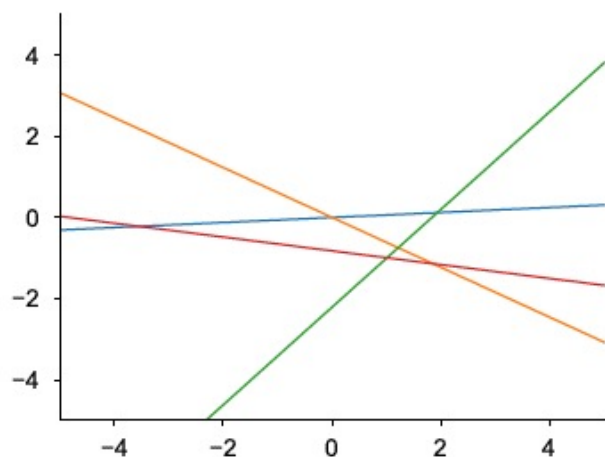
- 関数：入力 $x \mapsto y$ への写像
- 時間 t の関数としてみれば、軌跡
- 数学的には、関数解析で扱われる対象

代表的な使い方: ガウス過程回帰

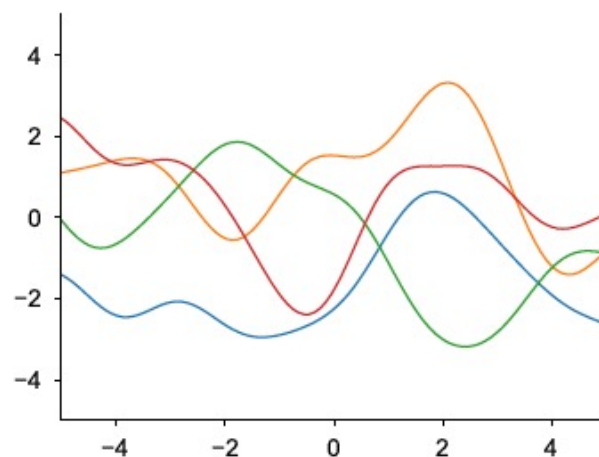
- (x,y) のペアが与えられた時、新しい x^* に対して y を予測
 - 非線形の、非常に柔軟な回帰が可能



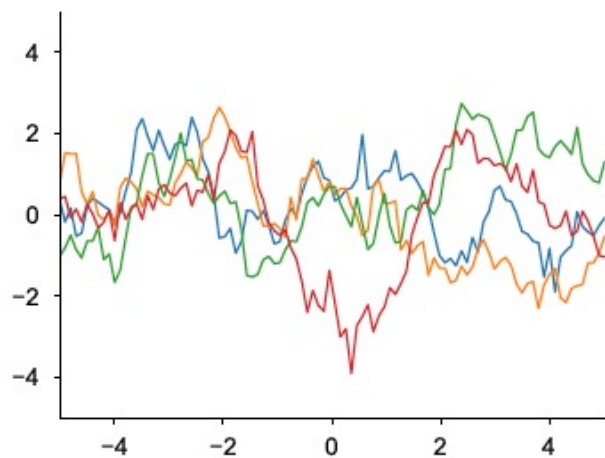
カーネルによる関数の違い (カーネルとは?)



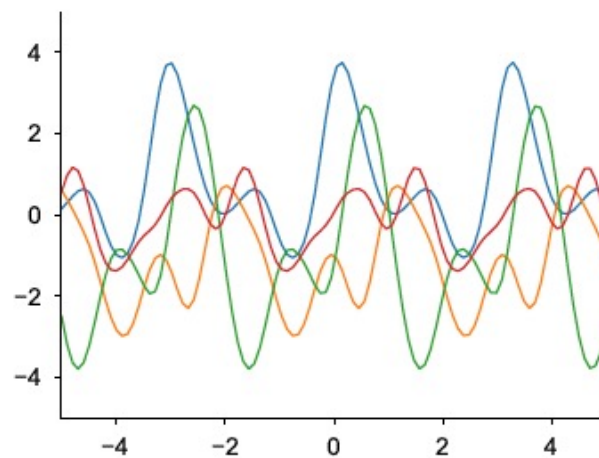
(a) 線形カーネル: $\mathbf{x}^T \mathbf{x}'$



(b) ガウスカーネル: $\exp(-|\mathbf{x} - \mathbf{x}'|^2/\theta)$



(c) 指数カーネル: $\exp(-|\mathbf{x} - \mathbf{x}'|/\theta)$
(Ornstein-Uhlenbeck 過程)



(d) 周期カーネル: $\exp(\theta_1 \cos(|\mathbf{x} - \mathbf{x}'|/\theta_2))$

カーネル: 入力の類似度を定義する関数

- 線形カーネル: $k(\mathbf{x}, \mathbf{x}') = \mathbf{x}^T \mathbf{x}'$
 - $\phi(\mathbf{x}) = \mathbf{x}$ を意味する → ガウス過程は、重回帰を包含
- 指数カーネル: $k(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{|\mathbf{x} - \mathbf{x}'|}{\theta}\right)$
- 周期カーネル: $k(\mathbf{x}, \mathbf{x}') = \exp\left(\cos \theta_1 \left(\frac{|\mathbf{x} - \mathbf{x}'|}{\theta_2}\right)\right)$
- Matérnカーネル:
$$k(\mathbf{x}, \mathbf{x}') = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu}r}{\theta}\right)^\nu K_\nu\left(\frac{\sqrt{2\nu}r}{\theta}\right) \quad (r = |\mathbf{x} - \mathbf{x}'|)$$
 - $\nu = \frac{1}{2}, \frac{3}{2}, \frac{5}{2}$ のときが有名 (Matérn 1/3/5)

GPLVM: 非線形な主成分分析

- Y だけが観測されているとき、

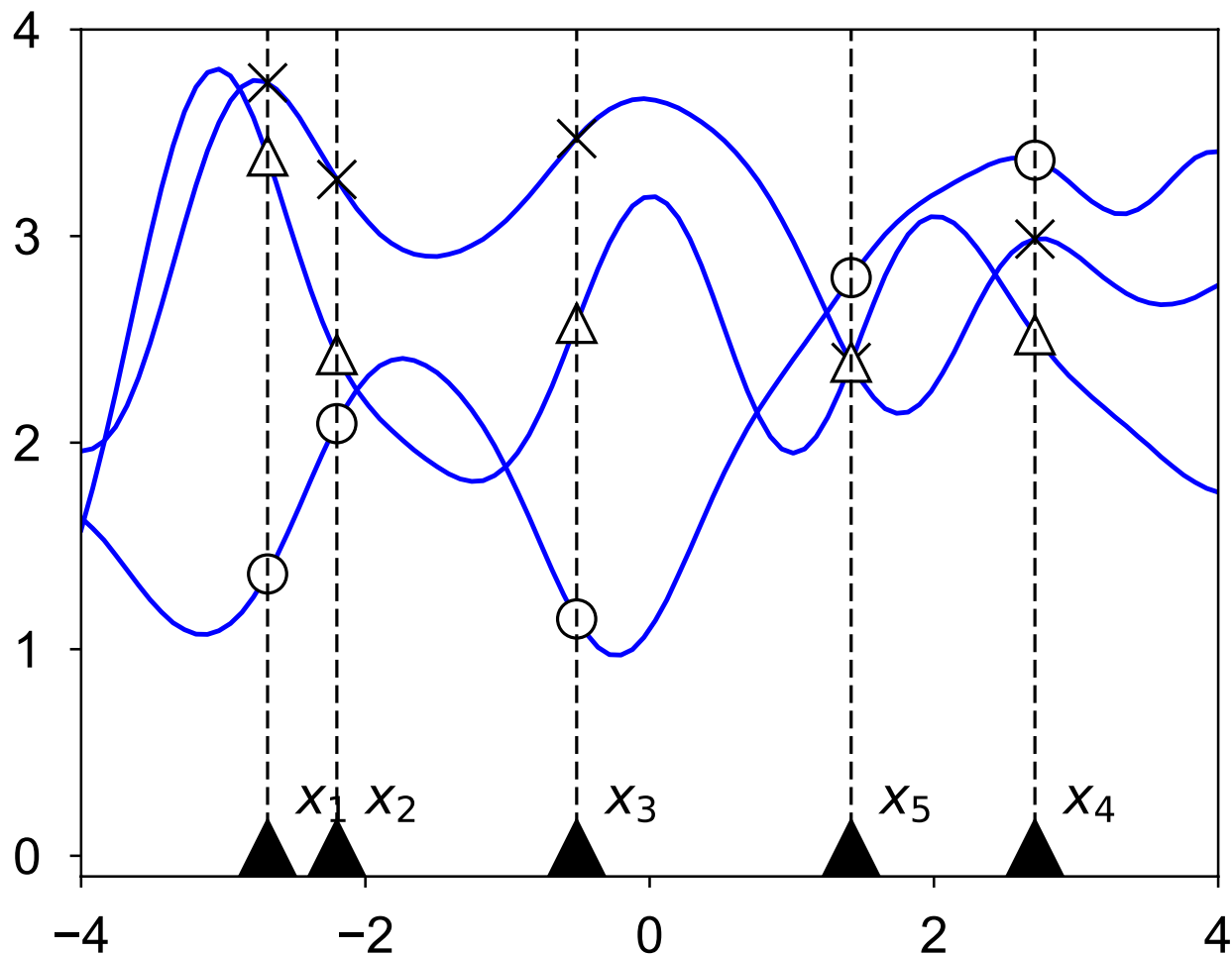
$$\log p(\mathbf{Y}|\mathbf{X}) = -\frac{DN}{2} \log(2\pi) - \frac{D}{2} \log |\mathbf{K}_{\mathbf{X}}| - \frac{1}{2} \text{tr}(\mathbf{K}_{\mathbf{X}}^{-1} \mathbf{Y}\mathbf{Y}^T)$$

$$\mathbf{K}_{\mathbf{X}} = \alpha \mathbf{X}\mathbf{X}^T + \beta^{-1} \mathbf{I}$$

$$\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_N)$$

- これを最大化する潜在的な X を見つければよい
- Gaussian Process Latent Variable Model (GPLVM)という (Lawrence+, NIPS 2003)

GPLVMのイメージ



- x が1次元の場合: 各観測値 (\times, Δ, \circ)の背後に x が存在

GPLVMの最適化

$$\log p(\mathbf{Y}|\mathbf{X}) = -\frac{DN}{2} \log(2\pi) - \frac{D}{2} \log |\mathbf{K}_{\mathbf{X}}| - \frac{1}{2} \text{tr}(\mathbf{K}_{\mathbf{X}}^{-1} \mathbf{Y}\mathbf{Y}^T)$$

$$\mathbf{K}_{\mathbf{X}} = \alpha \mathbf{X}\mathbf{X}^T + \beta^{-1} \mathbf{I}$$

$$\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_N)$$

- 自然にカーネル化されている → 任意のカーネルを導入

$$k(\mathbf{x}_n, \mathbf{x}_m) = \alpha \exp(-\gamma |\mathbf{x}_n - \mathbf{x}_m|^2) + \delta(n, m) \beta^{-1}$$

(RBF)

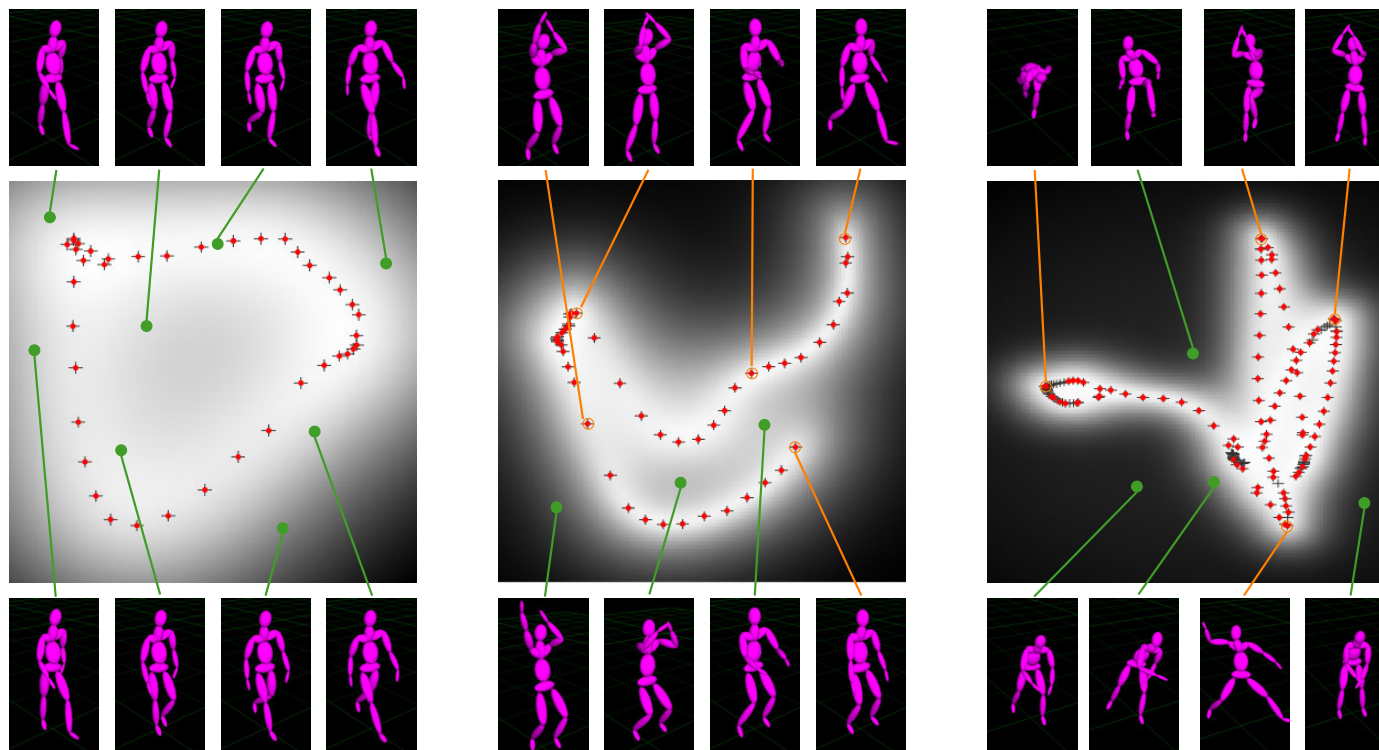
- $\frac{\partial L}{\partial \mathbf{K}_{\mathbf{X}}} = \mathbf{K}_{\mathbf{X}}^{-1} \mathbf{Y}\mathbf{Y}^T \mathbf{K}_{\mathbf{X}}^{-1} - D \mathbf{K}_{\mathbf{X}}^{-1}$

$$-\frac{\partial L}{\partial x_{nj}} = \frac{\partial L}{\partial \mathbf{K}_{\mathbf{X}}} \frac{\partial \mathbf{K}_{\mathbf{X}}}{\partial x_{nj}} \quad \text{を適用して微分}$$

– $\partial \mathbf{K}_{\mathbf{X}} / \partial x_{nj}$ は、上のカーネルから計算できる

GPLVM: 計算例

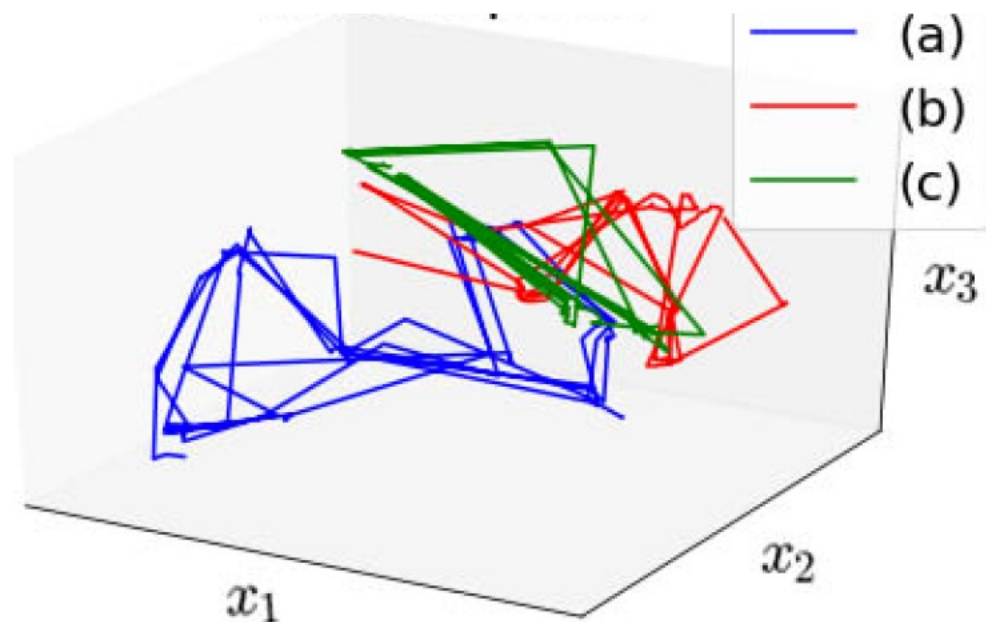
- “Style-based inverse kinematics”, SIGGRAPH 04



- 高次元の関節角ベクトルの時系列を、たった2次元程度に**非線形に**圧縮できる

運動のGPLVMの結果

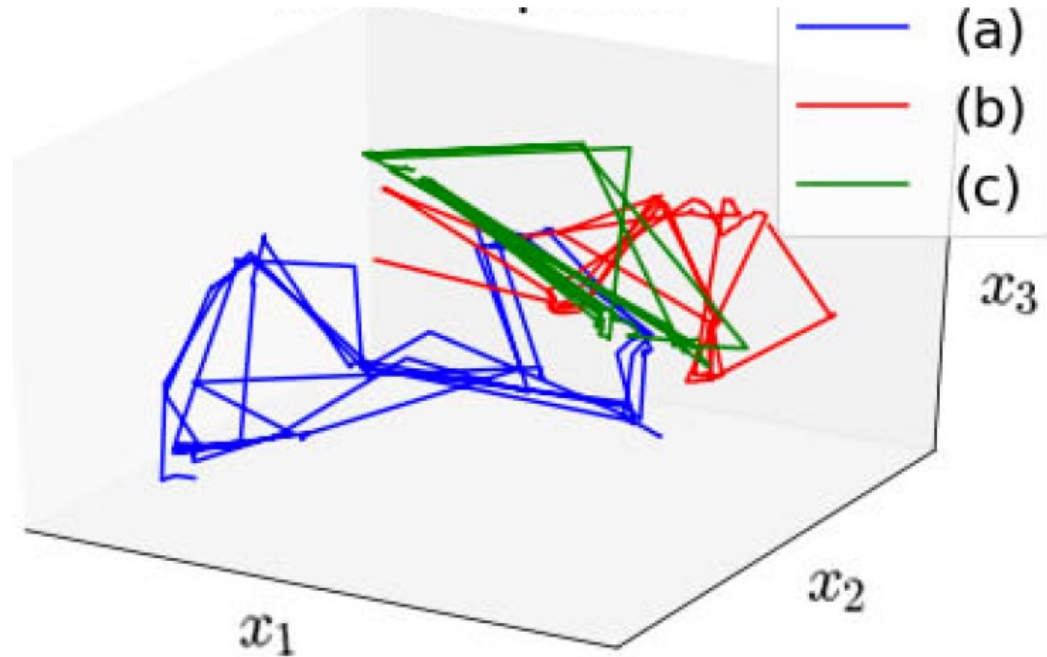
- (a)(b)(c)の異なる歩行動画について、GPLVMで軌跡を3次元に圧縮した結果



- 普通の主成分分析(線形)では、こんなに綺麗には圧縮できない

問題2：

- 「運動」の特徴をどうやってとらえるか？

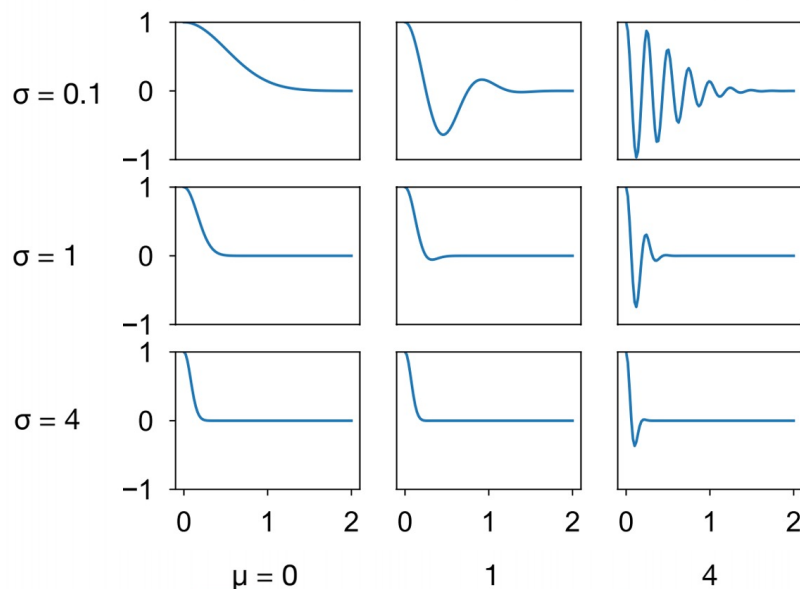


- “軌跡がどこを通るか”は重要ではない
- 「どういう関数か」 (=関数としてどんな種類に属するか) が重要

スペクトル混合カーネル [Wilson+ 2013]

- ガウス過程で使用するカーネル関数を,
データから自動的に学習する確率モデル

$$k(\tau) = \sum_{q=1}^Q w_q \cos(2\pi\tau^T \mu_q) \prod_{d=1}^D \exp(-2\pi^2 \tau_d^2 v_q^d)$$



基底関数

スペクトル混合カーネルとは

- ガウス過程のカーネルとして、値が $\tau = x - x'$ だけに依存する、定常カーネル $k(\tau)$ を考える
- ボホナーの定理により、任意の $k(\tau)$ は

$$k(\tau) = \int_{\mathbb{R}^D} e^{2\pi i s^T \tau} \psi(ds)$$

の形に表せる (逆フーリエ変換)

– $\psi(s)$ が、周波数領域での $k(\tau)$ の等価な表現



$k(\tau)$ を確率分布で表せる！

理論 (1)

- 通常の高スカーネル

$$k(x, x') = \exp\left(-\frac{1}{2}(x - x')^2 / \ell^2\right)$$

の周波数表現は、

$$S(s) = (2\pi\ell^2)^{1/2} \exp(-2\pi^2\ell^2 s^2)$$

– 中心0, 分散 $1/\ell^2$ の高ス分布!

理論 (2)

- $k(\tau)$ は周波数領域での確率密度 $\psi(s)$ と等価なので、 $\psi(s)$ に関して混合ガウス分布を考える
 - 0に関して対称なので、正だけ考えて鏡映

$$\phi(s | \mu, \sigma^2) = \mathcal{N}(s | \mu, \sigma^2)$$

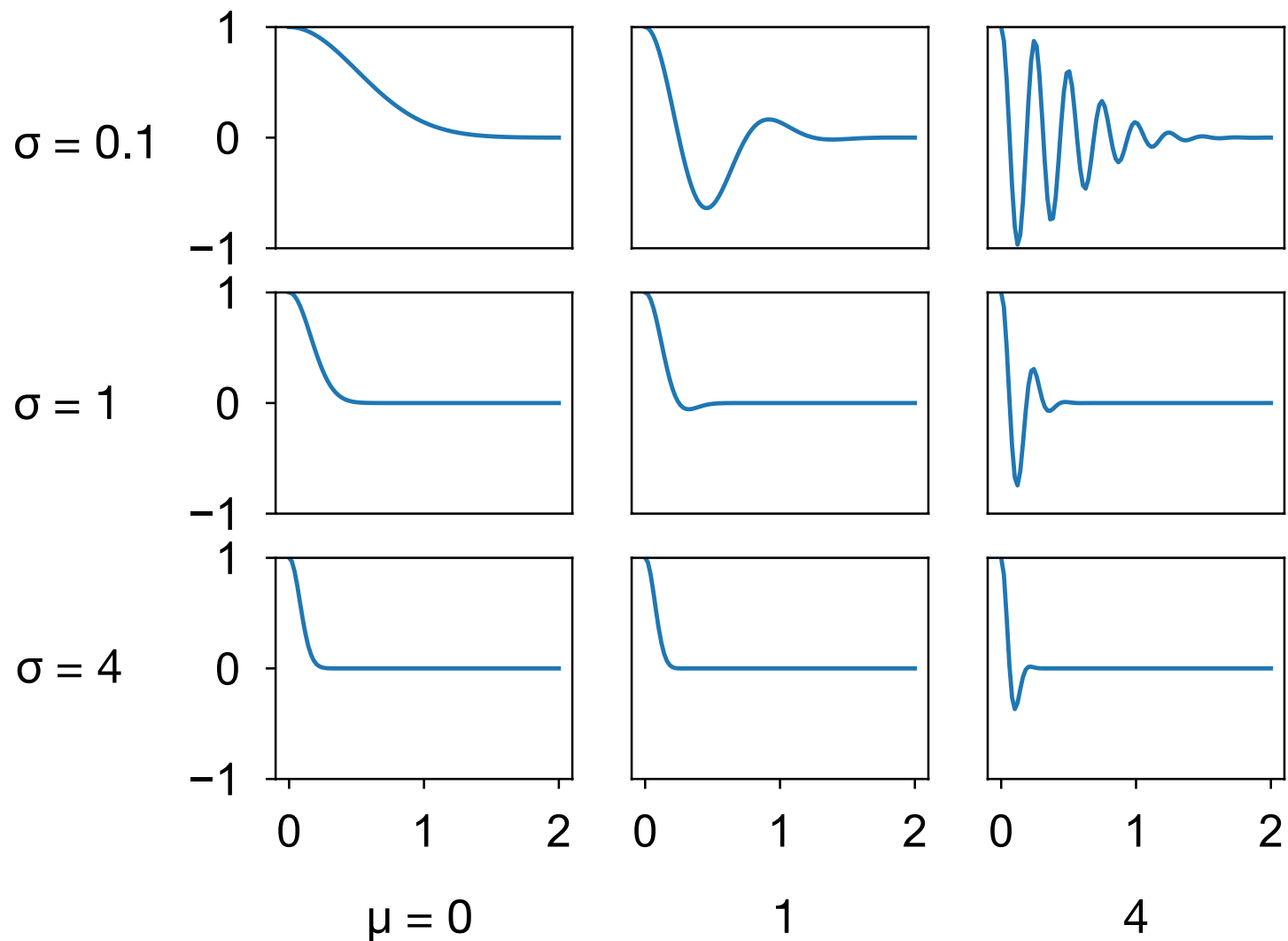
$$S(s) = (\phi(s) + \phi(-s))/2$$

- ガウス分布の各要素は、もとの領域では以下のカーネル関数を考えていることと等価

$$k(\tau | \sigma, \mu) = \exp(-2\pi^2 \sigma^2 \tau^2) \cos(2\pi \mu \tau)$$

要素カーネル

$$k(\tau|\sigma, \mu) = \exp(-2\pi^2\sigma^2\tau^2) \cos(2\pi\mu\tau)$$



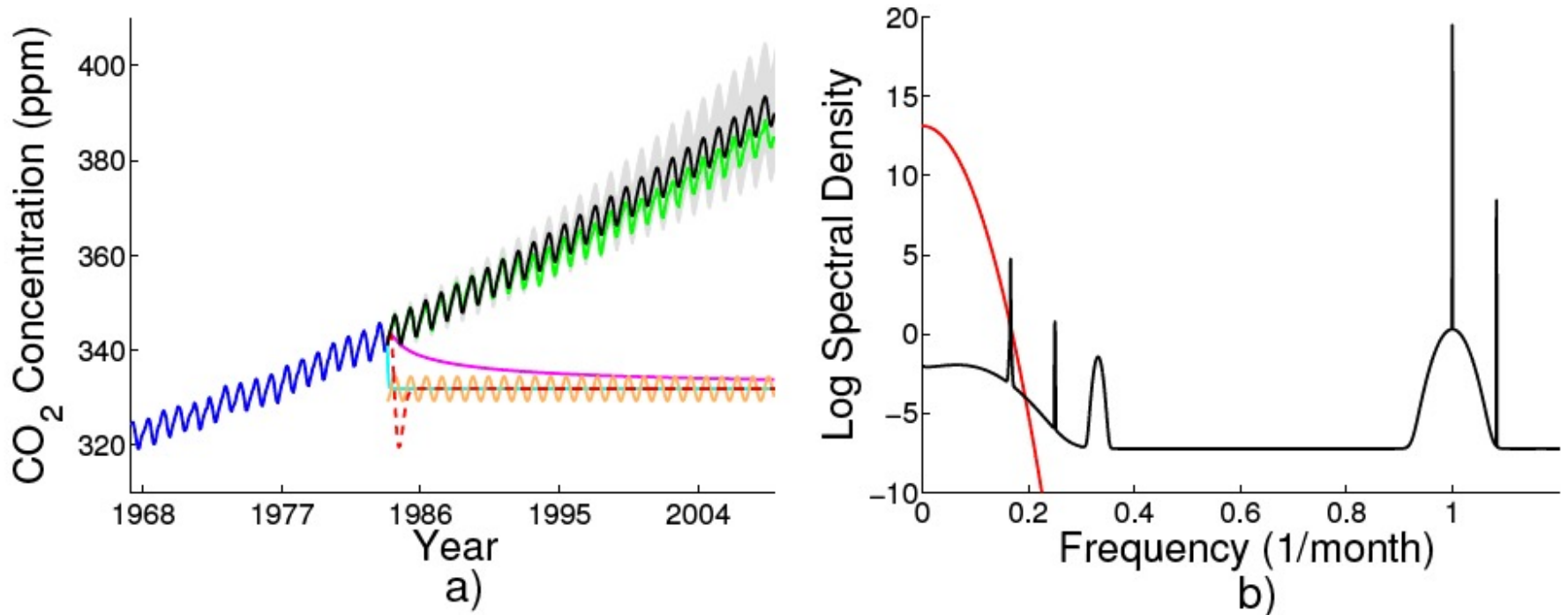
理論 (3)

- すなわち、これはカーネルとして、次の混合を考えていることになる (スペクトル混合カーネル)

$$k(\boldsymbol{\tau}) = \sum_{p=1}^P w_p \cos(2\pi \boldsymbol{\tau}^T \boldsymbol{\mu}^{(p)}) \exp\left(-\sum_{d=1}^D 2\pi^2 \sigma_d^{(p)} \tau_d^2\right)$$

- パラメータ w 、 μ 、 σ は通常のハイパーパラメータ最適化で学習できる
- ARD事前分布を使うことで、不要なガウス分布を除去している

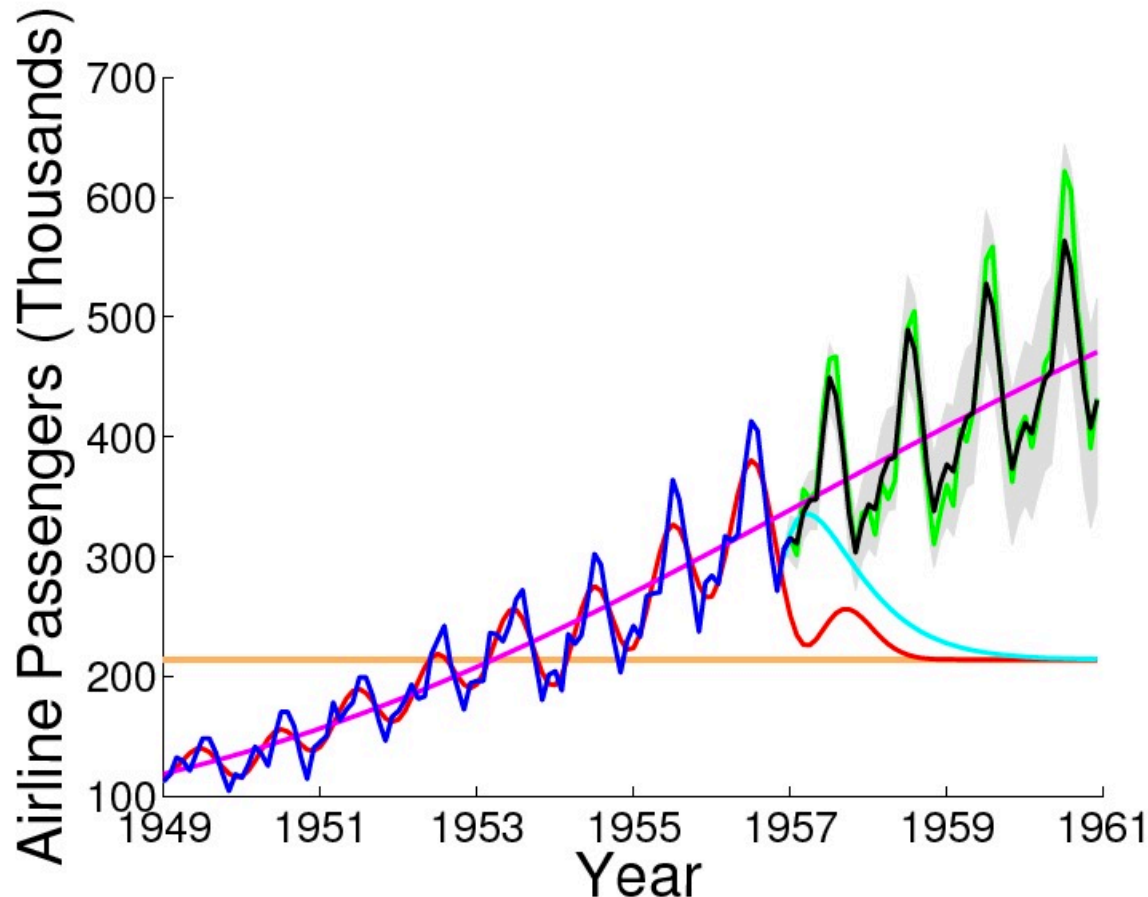
マウナロアCO₂濃度データ



- 青 = 学習データ、緑 = テストデータ
- 黒 = SM kernelによる予測
- 赤 = RBFカーネル

Airline Passengerデータ

- 1949-1961の毎月の航空乗客数のうち、最初の8年を学習に使って残りの4年分を予測



スペクトル混合カーネル→周波数抽出

- 入力：GPLVMで最適化した3次元の潜在変数
 - データを1次元ごとに分解してそれぞれ入力
- 出力：最適化された各カーネルの重み, 平均, 分散
- 混合するカーネルの個数は4個と10個に設定

$$k(\tau) = \sum_{q=1}^Q w_q \cos(2\pi\tau^T \mu_q) \prod_{d=1}^D \exp(-2\pi^2 \tau_d^2 v_q^d)$$

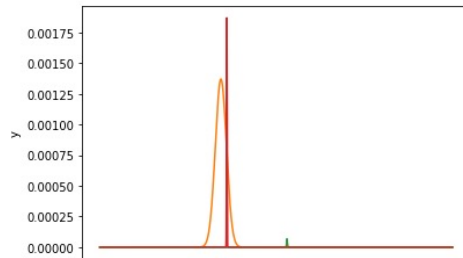
スペクトル混合カーネル→周波数抽出 (2)

- 学習した重み w 、平均 μ 、分散 v を用いて
ガウス分布を描画

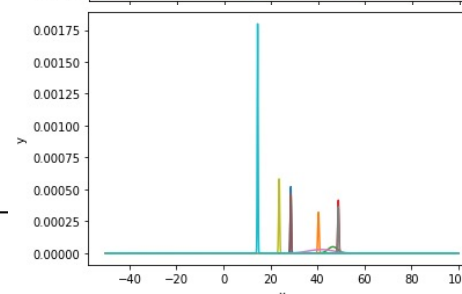
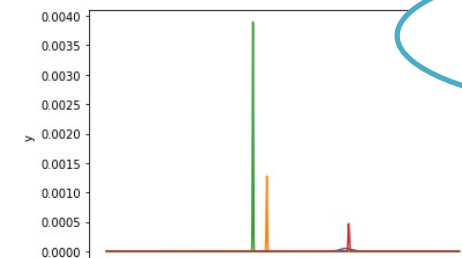
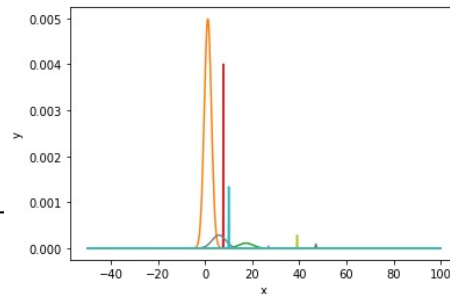
$$k(\tau) = \sum_{q=1}^Q w_q \cos(2\pi\tau^T \mu_q) \prod_{d=1}^D \exp(-2\pi^2 \tau_d^2 v_q^d)$$



Q=4



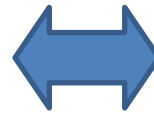
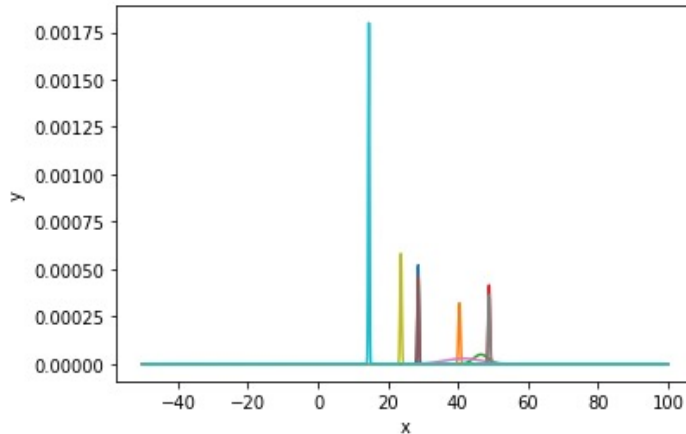
Q=10



Qを大きくすることで
細かい情報が追加

問題3：

- 動画の関数空間での周波数と、観測された副詞をどうやって結びつけるか？



自信を持って、足早に、
颯爽と、爽快に、キビ
キビと、しなやかに

周波数

副詞

Latent Dirichlet Allocation (LDA) (Blei+2003)

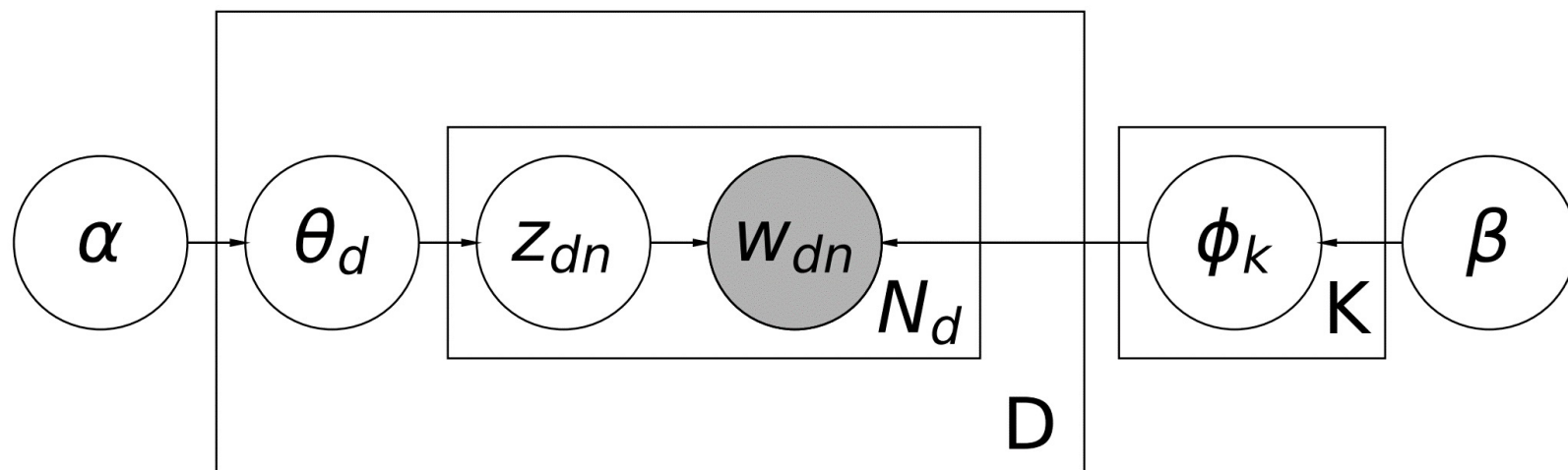
- テキストの各単語に、それを生成した潜在トピックを割り当てる確率モデル (Blei+2003より)

The William Randolph Hearst Foundation will give \$1.25 million to Lincoln Center, Metropolitan Opera Co., New York Philharmonic and Juilliard School. “Our board felt that we had a real opportunity to make a mark on the future of the performing arts with these grants an act every bit as important as our traditional areas of support in health, medical research, education and the social services,” Hearst Foundation President Randolph A. Hearst said Monday in announcing the grants. Lincoln Center’s share will be \$200,000 for its new building, which will house young artists and provide new public facilities. The Metropolitan Opera Co. and New York Philharmonic will receive \$400,000 each. The Juilliard School, where music and the performing arts are taught, will get \$250,000. The Hearst Foundation, a leading supporter of the Lincoln Center Consolidated Corporate Fund, will make its usual annual \$100,000 donation, too.

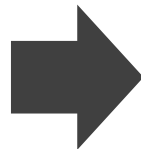
- 単語の色 = 潜在トピック
- 人間が教えなくても、完全に自動的に推定できる

LDAの確率モデル

- テキストはK個の潜在トピックを割合 θ で混合したものと仮定
 - 各単語 w にトピック z を割り当てていく
- θ および、トピック z から単語 w が出現する確率分布 $\{p(w|z)\}$ はディリクレ分布から生成されたと仮定する



動画一副詞の場合

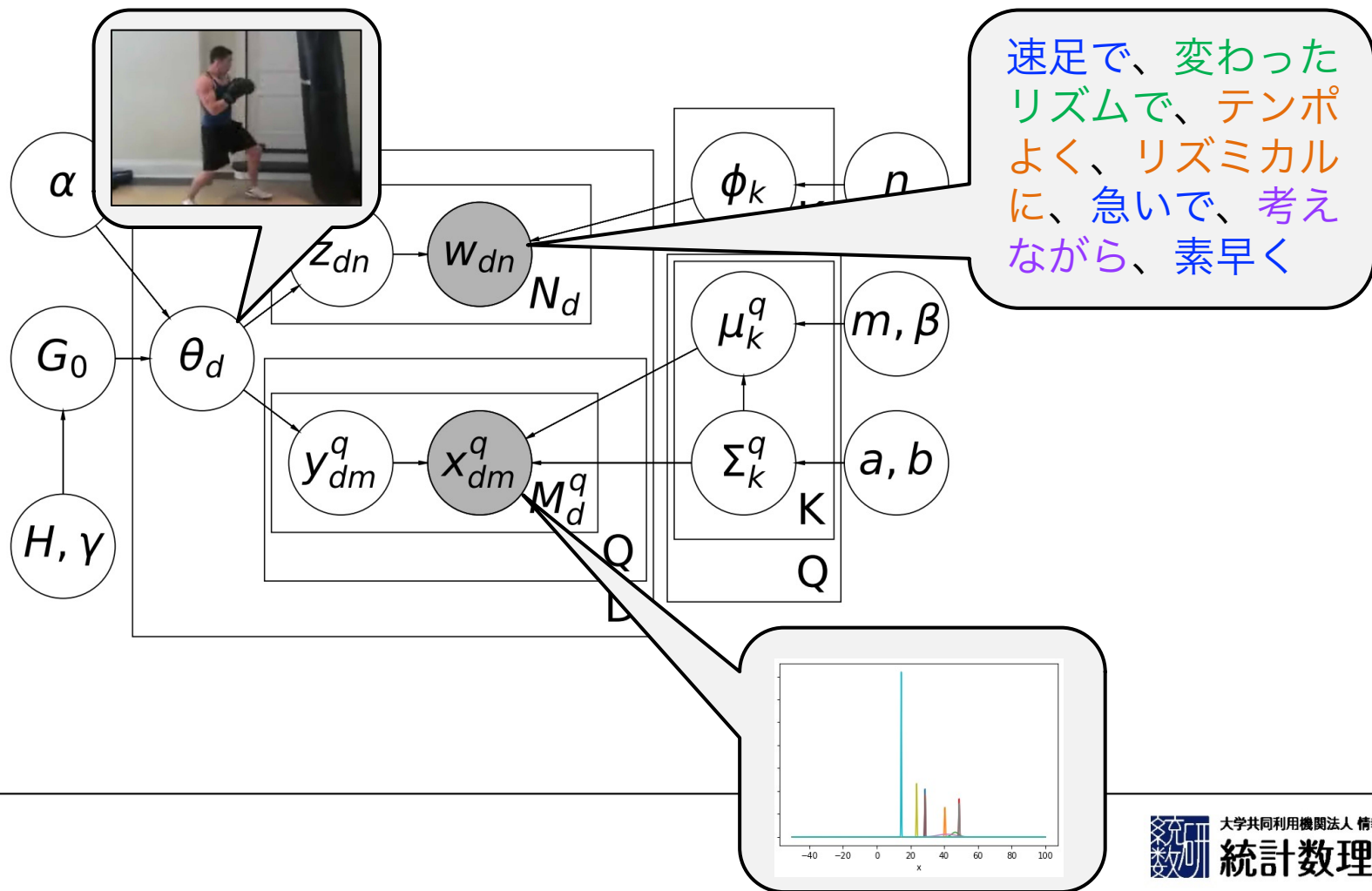


速足で、変わったリズムで、
テンポよく、リズムカルに、
急いで、考えながら、素早く

- 各動画には、副詞がクラウドでアノテーションされている
- LDAを使えば、それぞれの副詞の潜在トピック(=色)がわかる
- 動画には本当は、動作のカーネル関数空間での周波数情報がある！

HDP Spectral Mixture LDA

- 副詞と周波数を同時に潜在トピックでモデル化
- トピック数 K も自動推定 (HDP: 階層ディリクレ過程)



HDP-SMLDA 実験概要

- 入力：副詞データ，周波数データ
- データセット：100 Walks 95動画, AIST++ 1063動画
- 基底数：4, 10
- 集中度 α と単語分布のパラメータ β を最適化
- MCMCの繰り返し数：1000
 - 単語のパープレキシティを計算
 - スペクトル混合カーネル推定の際の重み W を用いて繰り返しごとに周波数をサンプリング

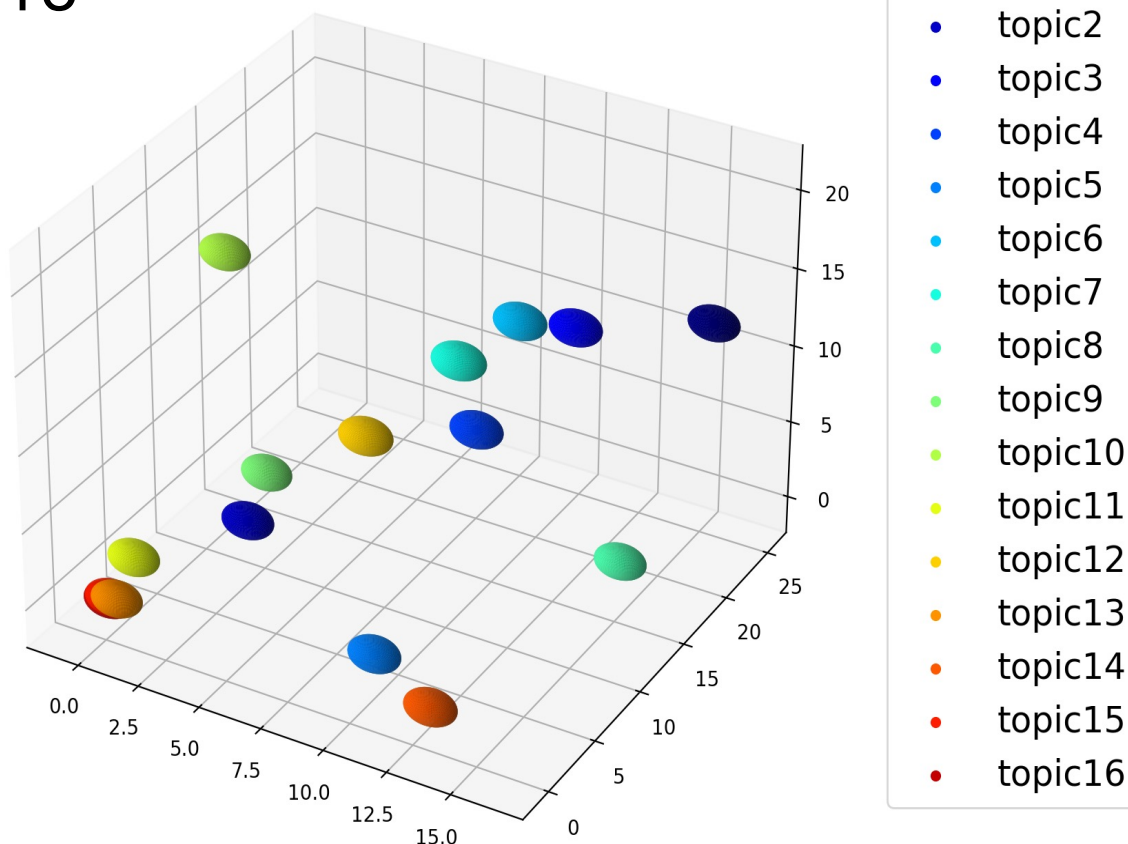
$$\text{perplexity}(\mathbf{w}_{\text{test}}) = \exp \left(- \frac{\sum_{d=1}^{D_{\text{test}}} \sum_{n=1}^{N_d} \log(p(w_{dn}))}{\sum_{d=1}^{D_{\text{test}}} N_d} \right)$$

HDP-SMLDA : AIST++データ

- 推定されたトピック数
 - Q=4 のとき K=16
 - Q=10 のとき K=16

各次元に関して
均等に配置

各トピックを象徴する
周波数の分布



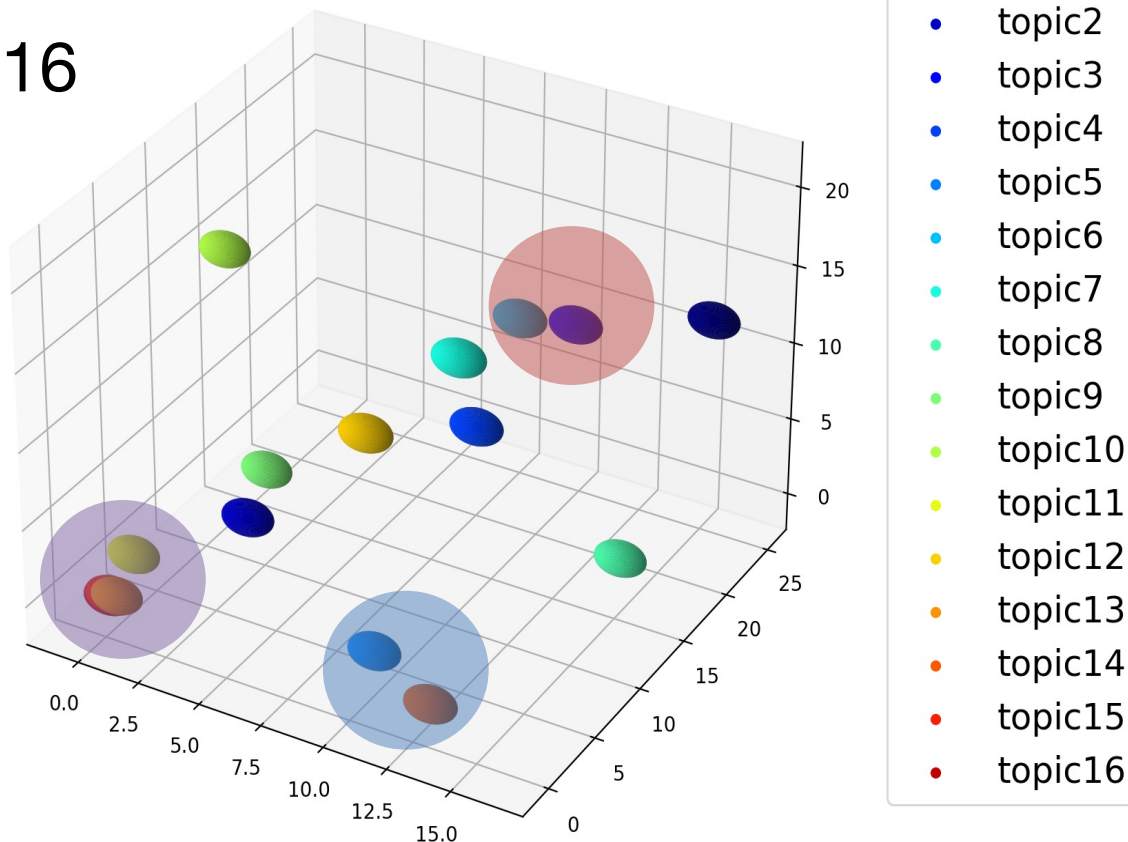
学習されたガウス分布の平均 μ (Q=4)

HDP-SMLDA : AIST++データ

- 推定されたトピック数
 - Q=4 のとき K=16
 - Q=10 のとき K=16

各次元に関して
均等に配置

各トピックを象徴する
周波数の分布



学習されたガウス分布の平均 μ (Q=4)

47

HDP-SMLDA 実験結果

AIST++データ Q=4 トピック別上位5語

topic: 1
激しく
力強く
はっきりと
熱心に
上品に

なめらかに
軽やかに
くるくると
topic: 5
力強く
激しい
激しく

topic: 2
楽しそうに
リズムカルに
軽やかに
弾むように
元気に

素早く
大胆に
topic: 6
踊るように
ステップを踏む
嬉しそうに
躍動するよう

topic: 3
規則正しく
テンポよく
躍動的に
生き生きと
大胆に

つまらなそうに
topic: 7
慣れたように
安定的に
くねくねと
キビキビと
ダイナミックに

topic: 4
しなやかに
優雅に

topic: 8
テンポ良く
スタイリッシュに
気持ち良さそうに
流れるように
格好良く

topic: 9
ゆったりと
滑らかに
ゆっくりと
機械的に
ゆるやかに

topic: 10
ダイナミックに
激しく
くねくねと
おおきく
キレよく

topic: 11
はずむように
ひろがるように

たどたどしく
ぐんぐんと
落ち着いた
topic: 12
格好よく
カクカクと
おおらかに

楽しそうな
機械のように
topic: 13
キビキビと
機械のように
コミカルに
しっかりと
ロボットのように

topic: 14
小刻みに
回るように
細かく
クルクルと
リズム感よく

topic: 15
確かめるように
ひょうきんに
丁寧に
慎重そうに
探すように
topic: 16

軽々しく
揺れているような
波のような
細かい動作で
ロボットのような

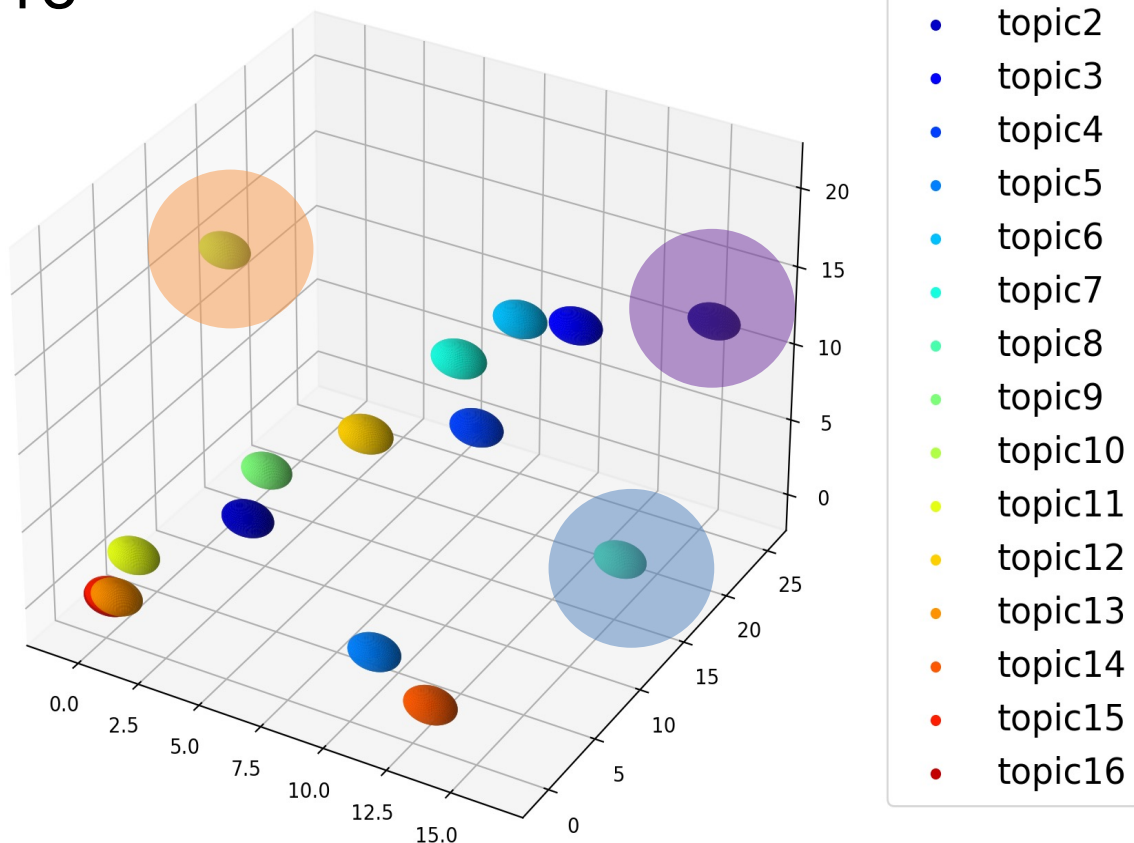
μ の距離が近い
トピック

HDP-SMLDA : AIST++データ

- 推定されたトピック数
 - Q=4 のとき K=16
 - Q=10 のとき K=16

各次元に関して
均等に配置

各トピックを象徴する
周波数の分布



学習されたガウス分布の平均 μ_i (Q=4)

HDP-SMLDA : AIST++データ

Q=4 トピック別上位5語

topic: 1

激しく
力強く
はっきりと
熱心に
上品に

topic: 2

楽しそうに
リズムカルに
軽やかに
弾むように
元気に

topic: 3

規則正しく
テンポよく
躍動的に
生き生きと
大胆に

topic: 4

しなやかに
優雅に

なめらかに
軽やかに
くるくると

topic: 5

力強く
激しい
激しく
素早く
大胆に

topic: 6

踊るように
ステップを踏む
嬉しそうに
躍動するように
つまらなそうに

topic: 7

慣れたように
安定的に

くねくねと

キビキビと
ダイナミックに

topic: 8

テンポ良く
スタイリッシュに
気持ち良さそうに
流れるように
格好良く

topic: 9

ゆったりと
滑らかに
ゆっくりと
機械的に
ゆるやかに

topic: 10

ダイナミックに
激しく
くねくねと
おおきく
キレイよく

topic: 11

はずむように
ひろがるように

たどたどしく
ぐんぐんと
落ち着いた

topic: 12

格好よく
カクカクと
おもしろかに

楽しそうな
機械のように

topic: 13

キビキビと
機械のように
コミカルに
しっかりと
ロボットのように

topic: 14

小刻みに
回るように
細かく
クルクルと
リズム感よく

topic: 15

確かめるように
ひょうきんに
丁寧に
慎重そうに
探すように

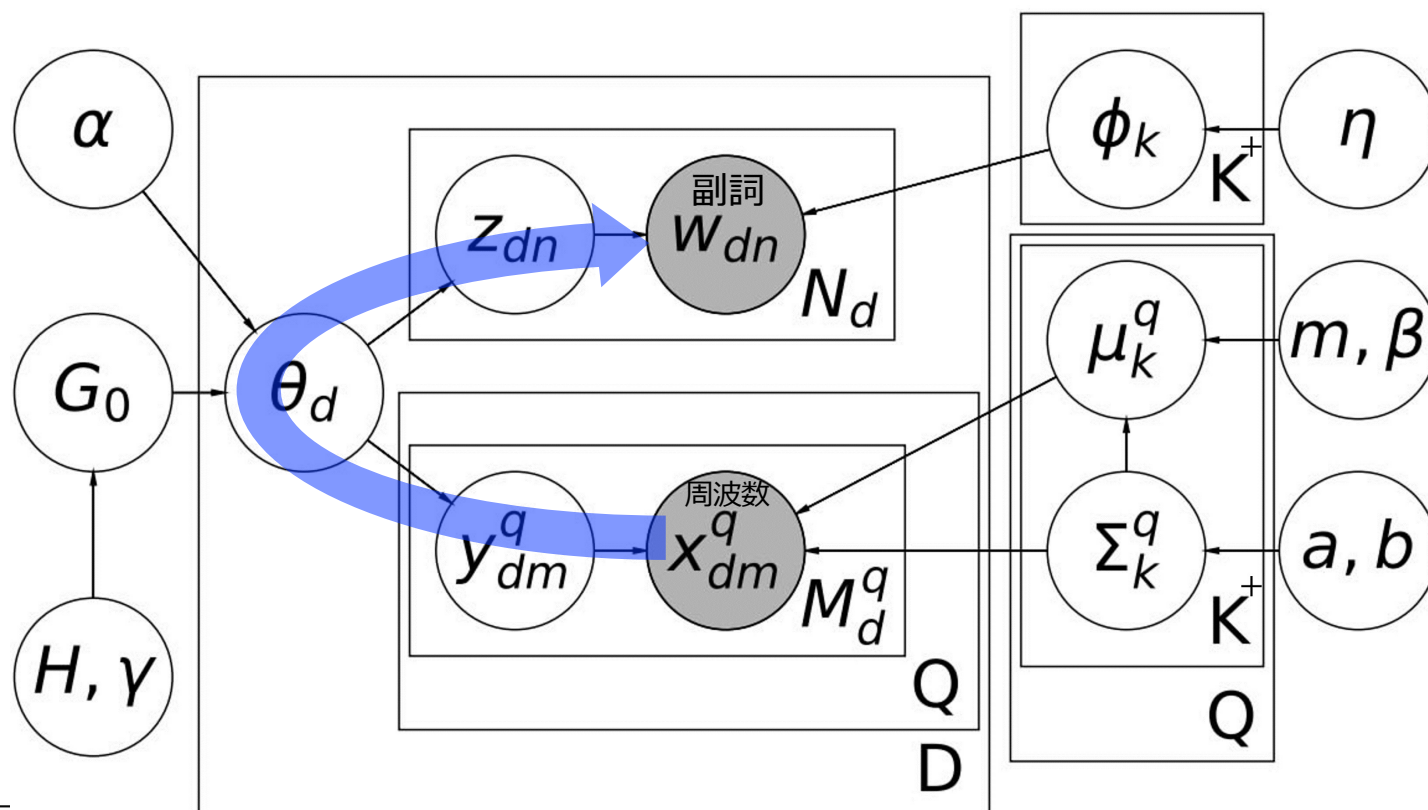
topic: 16

軽く
揺れているような
波のような
細かい動作で
ロボットのような

μ が特異な
トピック

周波数情報から副詞の生成

- 学習した単語分布 ϕ を用いて
評価動画の周波数情報から副詞を生成



周波数情報から副詞の生成 上位10語

実際に付与された副詞

明るい
魅せるように
恋するように
欲するように
大きな
軽やかに
爽快に
情熱的に
リズムカルに
陽気に
気持ち良さそうに
軽快に
スタイリッシュに
テンポ良く
格好良く
センス良く
スムーズに
流れるように
大胆に

モデルの予測 (Q=4)

力強く
激しい
激しく
素早く
大胆に
堂々と
キビキビと
ダイナミックに
パワフルに
かっこよく

モデルの予測 (Q=10)

テンポ良く
スムーズに
スタイリッシュに
流れるように
陽気に
悲しそうに
格好良く
気持ち良さそうに
センス良く
堂々と



各モデルにおけるPerplexity (= 1/予測確率)

- 100 Walks データ

LSTM / GPLVM	LSTM / Original	MLP (Q=4)	MLP (Q=10)	HDP-SMLDA (Q=4)	HDP-SMLDA (Q=10)
210	402	253	284	89	117

提案手法が
ベストスコア

- AIST++ データ

LSTM / GPLVM	LSTM / Original	MLP (Q=4)	MLP (Q=10)	HDP-SMLDA (Q=4)	HDP-SMLDA (Q=10)
1068	1794	994	1027	320	382

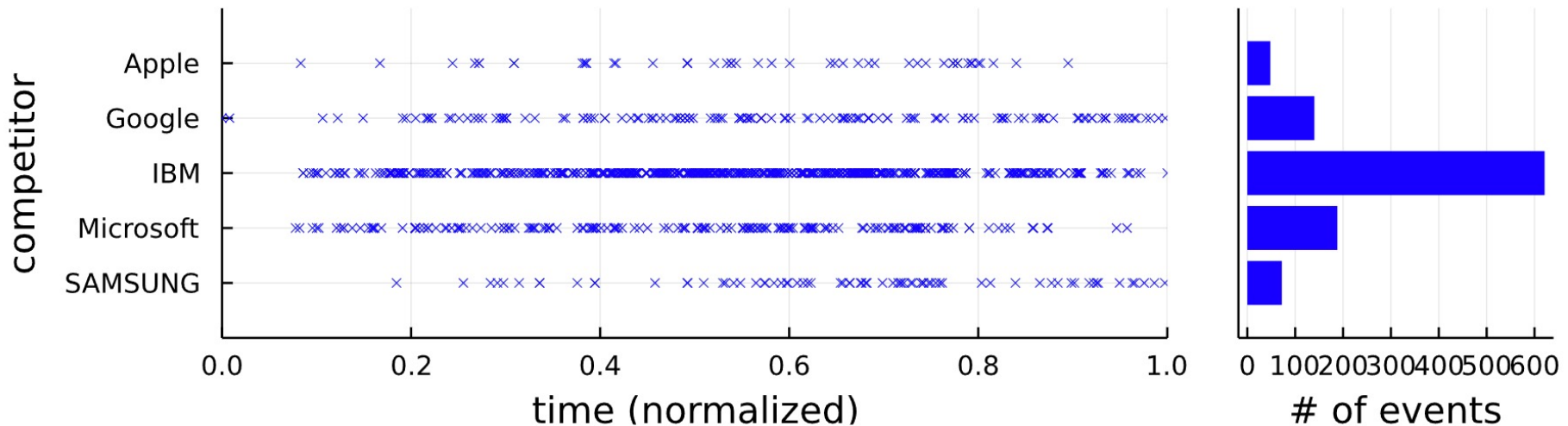
次元圧縮が
クラス分類に有効

離散データの背後にある 微分方程式の推定

(総研大統計科学専攻D3 宮澤脩一さん(AGC)との共同研究)

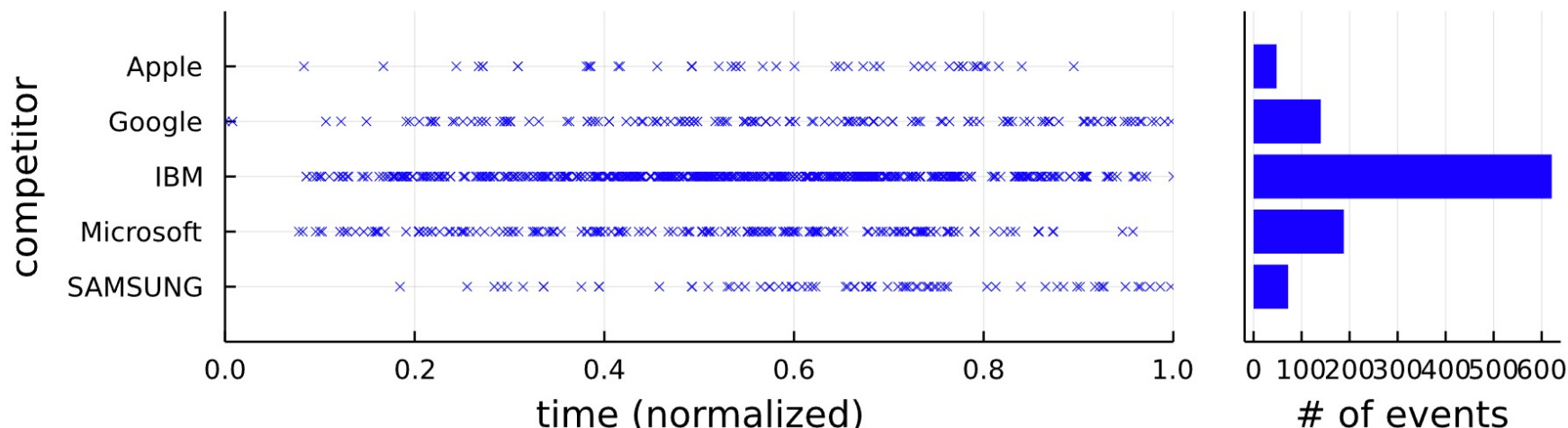
テキストの離散データの時系列

- 例：各社の特許出願イベントの時系列 (競争がある)



- 自然言語処理の“言語モデル”分野の特許データ
- テキストを潜在トピックに分解すれば、一般的なテキストでも同様の時系列が得られる

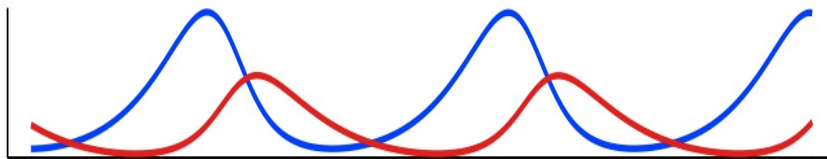
微分方程式？



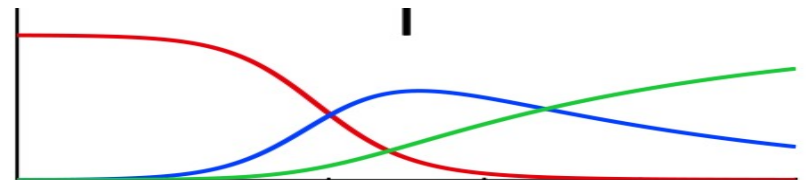
- こうしたデータは、何かの微分方程式に従っている可能性が高い
- しかし、観測値は離散 (-63.57のような連続値ではない!)
 - 統計的には、点過程とよばれる問題
 - 連続値ではないので、ルンゲクッタ法等は使えない!

常微分方程式 (ODE)

- 時間変化する複数の量の関係を記述する
 - 土木工学、物理学、電気工学、経済学、生物学...



捕食者-被捕食者



感染症のモデル (COVIDなど)

- 観測値を y 、潜在状態を x 、ガウスノイズを ϵ とすると

$$y(t_i) = x(t_i) + \epsilon(t_i), \quad i = 1, \dots, N.$$

このとき、 x の時間微分がある関数に従う

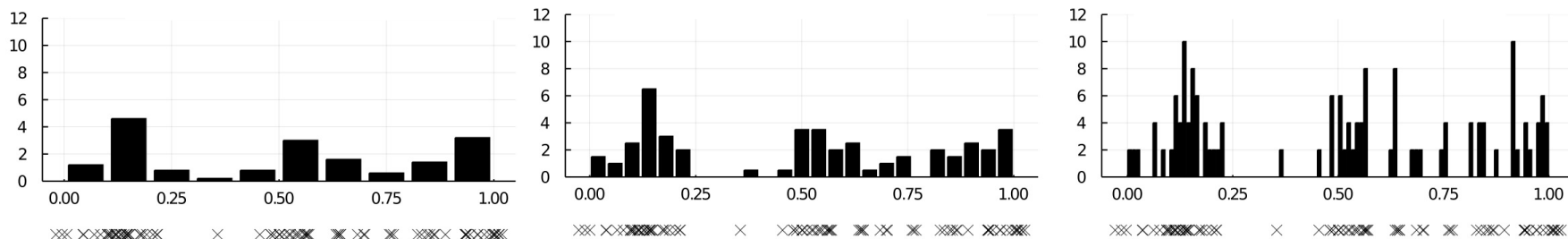
$$\dot{\mathbf{x}}(t) \equiv \frac{d}{dt}\mathbf{x}(t) = \mathbf{f}(\mathbf{x}(t), \boldsymbol{\theta}).$$

点過程データの取り扱い

離散化

粗

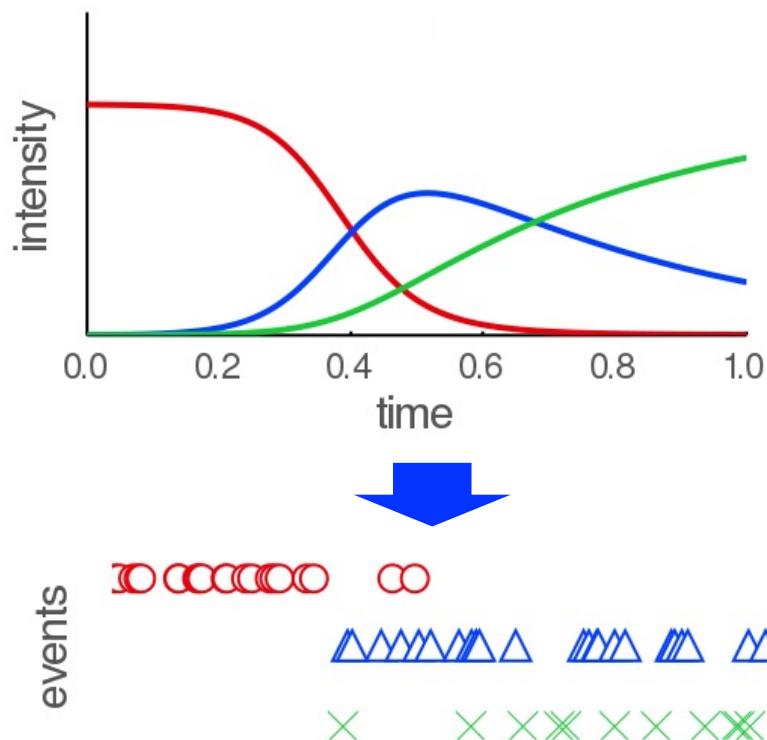
細



- 一定間隔で時間を離散化してヒストグラムにすれば、
一応は数値データにできる
↓
- しかし、
 - 粗い離散化：時間的な依存関係を丸めてしまう
 - 細かい離散化：頻度0のビンの増加、信頼度の低下、ガウスノイズの仮定が不適切に

Cox過程 (Cox, 1955)

- 点が観測される潜在的な強度が時間変化する確率モデル

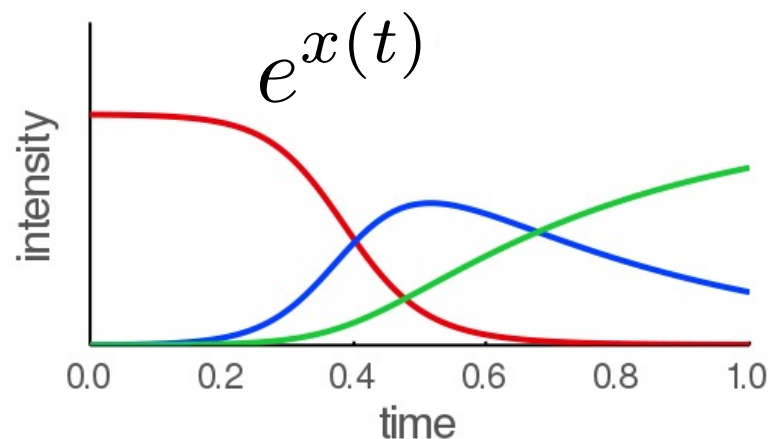


$$\lambda(t) = \lambda_0 \cdot \exp(x(t)) (> 0),$$
$$x(t) \sim GP(0, K)$$

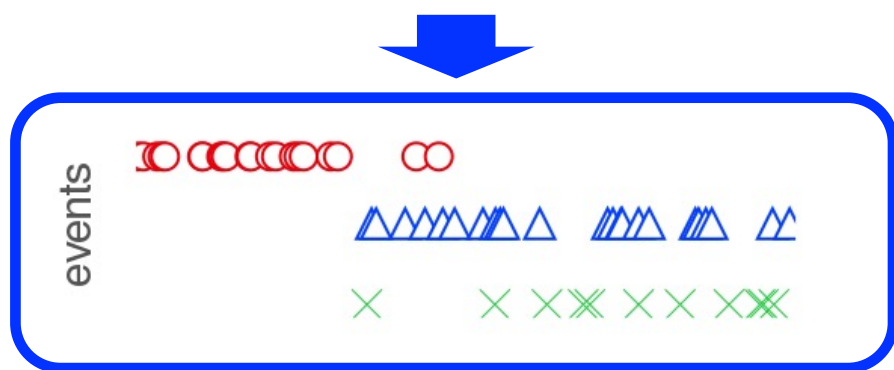
潜在的強度の対数がガウス過程に従う：
対数ガウスCox過程 (Møller+ 98)

提案法：ODE統合型Cox過程 (LGCP-GM)

- 潜在強度がランダムなガウス過程に従うのではなく、微分方程式(ODE)に従って時間発展するとする

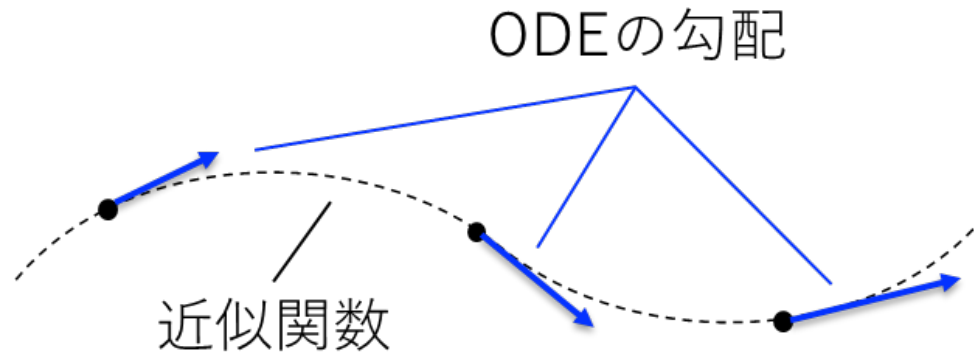


$$\frac{e^{x(t)}}{dt} = f(e^{x(t)}, \theta)$$



われわれが知っているデータはここだけ

問題: データが離散的



勾配マッチングの概念図

- 任意の時刻で観測データが得られるわけではない！
→ ガウス過程によって補間
- 補間したガウス過程の勾配を、推定したい微分方程式の勾配と同じにする
→ 勾配マッチング (Dondelinger+13, Wenk+19)

学習

- 観測値 \mathbf{y} 、潜在的強度 \mathbf{x} , ODEパラメータ θ の事後分布

$$p(\mathbf{y}, \mathbf{x}, \theta | \mathbf{m}, \sigma, \phi, \gamma) \propto p(\mathbf{m} | \mathbf{y}) \times p(\theta) \times \mathcal{N}(\mathbf{y} | \hat{\mathbf{x}}, \text{diag}(\hat{\mathbf{C}}_\phi) + \sigma^2 \mathbf{I}) \\ \times \mathcal{N}(\mathbf{x} | \mathbf{0}, \mathbf{C}_\phi) \times \mathcal{N}(\mathbf{g}(\mathbf{x}, \theta) | \mathbf{D}\mathbf{x}, \mathbf{A} + \gamma \mathbf{I}).$$

- 解析的には解けないので、MCMC法でサンプリングによって推定
 - Hamiltonian Monte Carlo+Blocked Gibbs

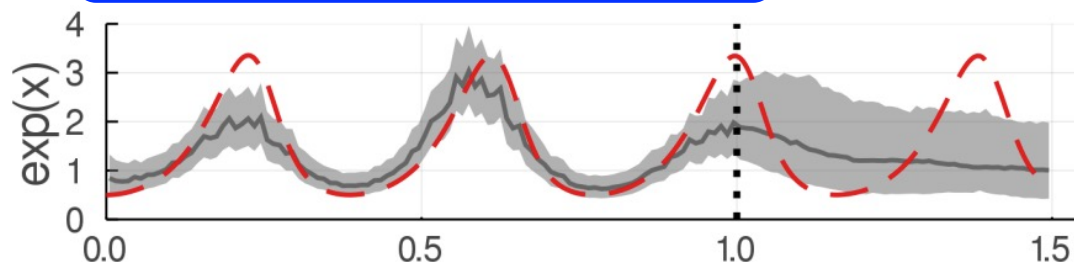
実験結果 (予測)

観測値はこれだけ

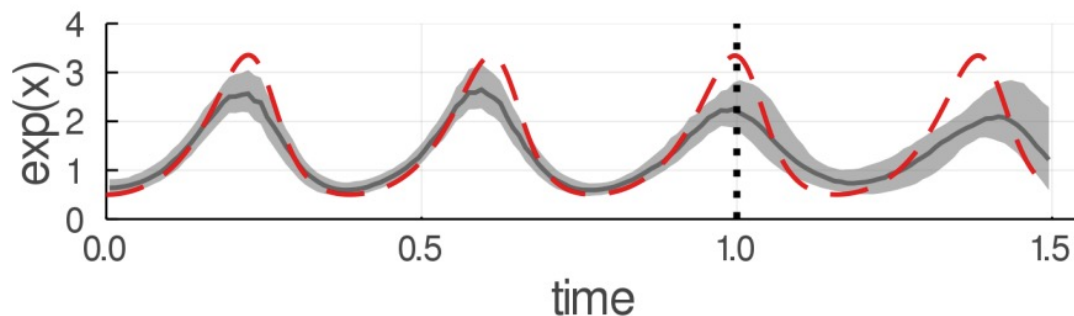
Observations



LGCP



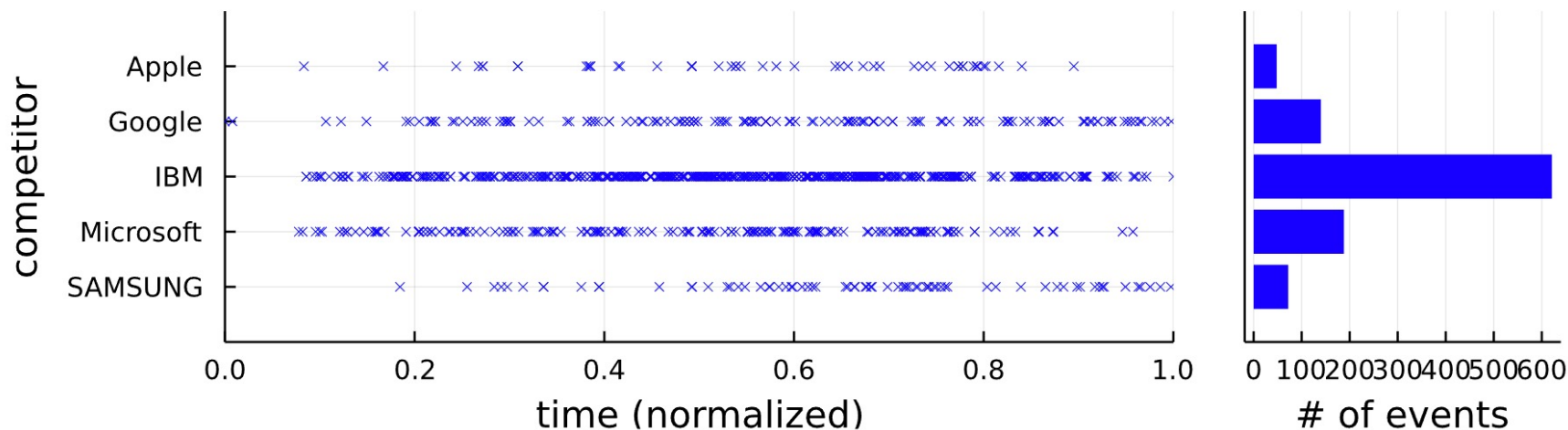
LGCP-GM
(proposed)



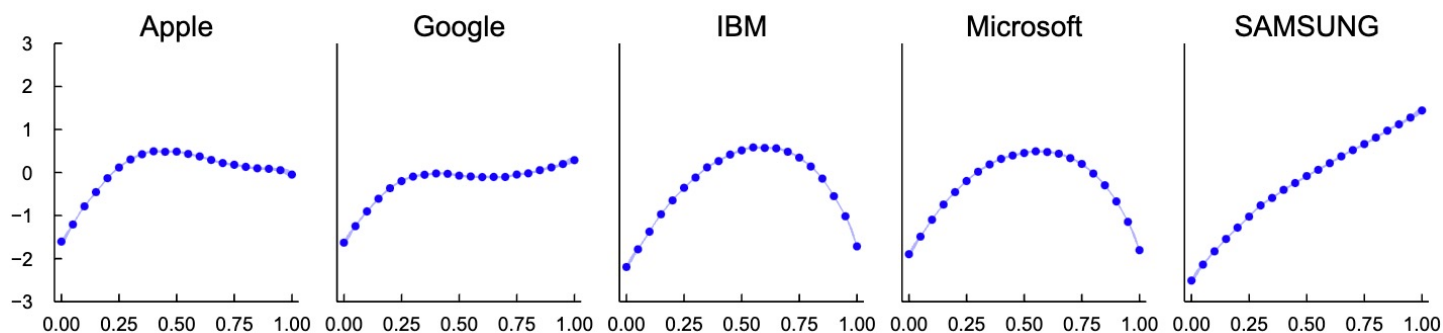
- 微分方程式を考慮しない対数ガウスCox過程に比べ、提案法は潜在ダイナミクスを考慮して高精度に予測できる

実験結果 (特許)

- 自然言語処理に関する特許の出願データ

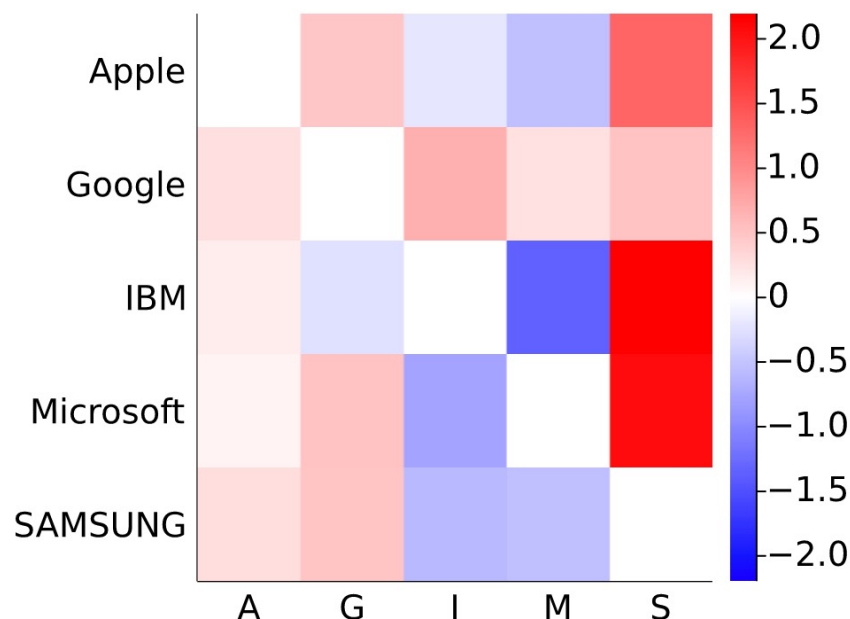


- 潜在強度 x の推定結果



実験結果 (特許)

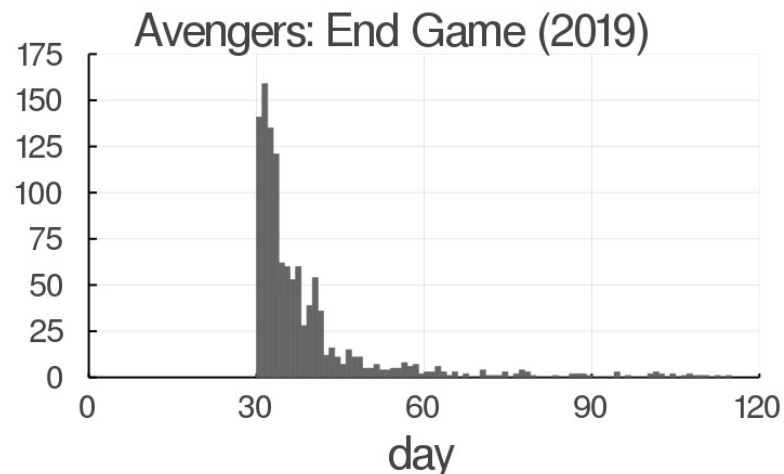
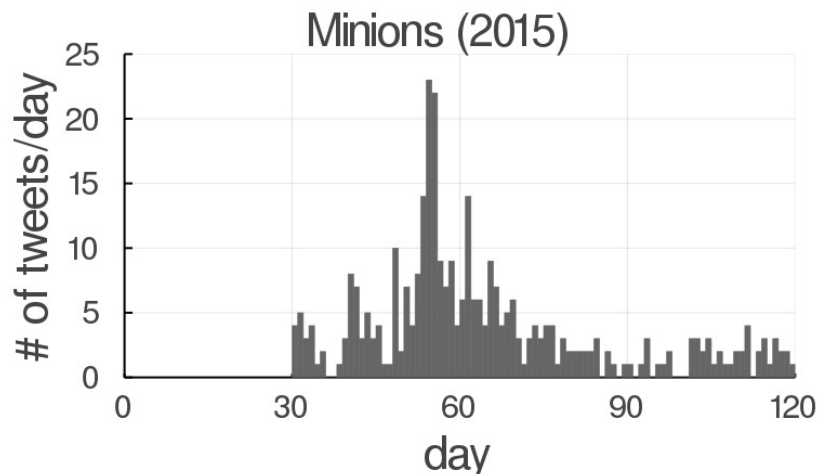
- 潜在強度 x の裏にある、Lotka-Volterra競争方程式のパラメータの推定値



- IBM→サムスン(+), IBM→Microsoft(-)、
Microsoft→サムスン (+) の影響関係がみられる

実験結果 (映画視聴)

MovieTweatings (Dooms+13)データセット



- 映画レビューのツイート数の時間変化
- 仮定する微分方程式：SIRモデル (“I”しか観測されない)
- パラメータの推定結果：

Movie Production	a	b
Minions (2015)	2.45 ± 0.737	8.71 ± 1.95
Avengers: End Game (2019)	4.95 ± 0.748	9.90 ± 0.149

a: trend up
b: trend down