脳デコーディングによる視覚像再構成の批判的再検討

長野 祥大 [1]
[1] 京都大学 大学院情報学研究科
e-mail : nagano@i.kyoto-u.ac.jp

1 概要

脳活動から任意の視覚体験を取り出すことを目指す再構成研究では、訓練データを超えた予測を含めた性能限界の正確な把握が極めて重要である。我々は近年注目を集める再構成手法で報告された結果が、訓練–テストデータ間の意味的 /視覚的類似性に大きく依存していることを示す。また、観察事実を元に構築した人工モデルを用いた一般的議論を試みる。訓練データが少数のクラスタの場合に予測可能な出力次元が制限され新規データを予測できない「出力次元崩壊」が起きることを示す。なお、本予稿および講演はおもにプレプリント [1] に基づく。

2 テキスト誘導型拡散モデルを用いた再構成手法の批判的再検討

視覚像再構成は、脳内表象に符号化された情報を特徴量として翻訳し、対応する視覚刺激として再構成する翻訳器-生成器パイプラインに基づく。翻訳器は脳活動と特徴量の組を用いた教師あり学習で得られ、生成器は予測された特徴量をもとに画像を生成する。DNN の階層的中間表現を予測特徴量とすることで、訓練データ外のカテゴリの画像が再構成できることが明らかになっている [2]。近年、画像のキャプションの DNN 表現（テキスト特徴量）をあわせて予測特徴量とし、生成器にテキスト誘導型拡散モデルを用いることでリアルな再構成画像が得られることが報告されている [3, 4]。しかし、先行研究と比較してこれらの手法の訓練データを超えた汎化性能は未知であった。再現実験では元の論文中で使用されたデータセット（NSD）では高品質な再構成画像が得られた一方、先行研究 [2] で用いられたデータセットの再構成は真の画像から逸脱していた。UMAP を用いた NSD のテキスト特徴量の可視化から訓練–テストデータ間の類似性が大きい 40 程度のクラスタ構造が確認されており、この結果は前述の高品質な再構成が訓練データへの過適合に由来することを示唆する。

3 混合ガウス分布を用いたモデル化と出力次元崩壊

混合ガウス分布を用いて前述の NSD のようなデータを模擬することで、クラスタ構造を持つ特徴量を予測することの問題点を明らかにする（図 1a)。特徴量 $\mathbf{y} \in \mathbb{R}^D$ を混合ガウス分布から生成し、教師重み $\bar{B} \in \mathbb{R}^{D\times D}$ とガウスノイズ $\boldsymbol{\epsilon}$ を用いて $\mathbf{x} = \bar{B}^\top \mathbf{y} + \boldsymbol{\epsilon}$ で脳活動 $\mathbf{x} \in \mathbb{R}^D$ を作成する。この訓練データを用いて脳活動 $\mathbf{x}$ から特徴量 $\mathbf{y}$ を視覚像再構成で広く用いられるリッジ線形回帰で予測する問題を考える。図 1b は訓練済み翻訳器がテストデータが属する真のクラスタを正しく同定できた精度を表し、オレンジ色の線はテストデータが訓練データに含まれない未知クラスタの場合を表す。訓練データセットが少数のクラスタしか含まない場合、たとえ訓練データ数が十分でも翻訳器は未知クラスタを全く同定できなかった。これは予測特徴量が真のテストクラスタよりも訓練クラスタのいずれかにより近かったことを意味する。対して訓練クラスタ数が一定の値を超えることで未知クラスタの同定が既存クラスタと同程度に可能になった。脳データ・特徴量の次元 D に対して未知クラスタの同定精度が十分に大きくなるのに必要な訓練クラスタ数を求めたところ、図 1c に示すよううに D に対しておおよそ線形オーダーであった。この結果は、訓練データの多様性が未知データへ

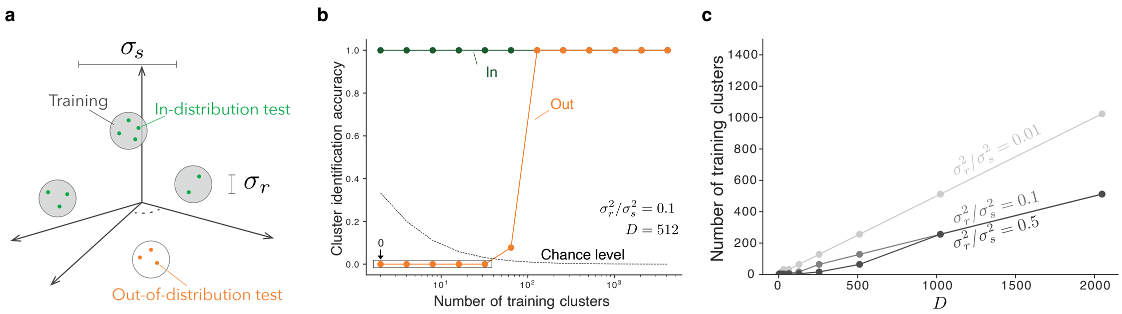

図 1: 人工データを用いた数値シミュレーション。**(a)**: 人工データの生成過程の模式図。**(b)**: 訓練クラスタ数に対するクラスタ同定精度。**(c)**: クラスタ同定に必要な訓練クラスタ数の次元依存性。

の汎化に寄与するが、空間全域を覆い尽くすほど大量のクラスタは不要であることを強調する。

次元 D データ数 P の訓練済み翻訳器の出力はレプリカ解析 [5] によって以下のように評価できる:

$$\hat{\mathbf{y}} = \mathbb{E}_{\mathcal{D}}[W_{\mathcal{D}}^*]^\top \mathbf{x} = \bar{A}^\top \operatorname{diag}\left(\frac{P\eta_\rho}{P\eta_\rho + \kappa}\right) \mathbf{\Phi}(\mathbf{x}) \tag{1}$$

ここで η_ρ は ρ 番目のデータの固有値、κ はリッジ正則化係数と訓練データから定まる定数で、$\mathbf{\Phi}(\mathbf{x})$ はカーネル関数であり、線形回帰では訓練データの共分散行列 $\mathbb{E}[\mathbf{x}\mathbf{x}^\top] = U\Sigma U^\top$ を用いて $\mathbf{\Phi}(\mathbf{x}) = \Sigma^{-1/2}U^\top\mathbf{x}$ と書ける。なおデータ生成に用いた教師重み $\bar{B}$ と上式の $\bar{A}$ の対応関係を $\bar{B} = \bar{W}^{-1} = \left(U\Sigma^{-1/2}\bar{A}\right)^{-1}$ とした。リッジ回帰で予測される特徴量は真のモデルの出力 $\mathbf{y} = \bar{A}^\top\mathbf{\Phi}(\mathbf{x})$ に対して、各固有方向が係数 $P\eta_\rho/(P\eta_\rho + \kappa)$ の分だけ歪むことがわかる。結果として、少ないクラスタ数などデータの多様性が低い状況ではデータ数 P と固有値 η_ρ の積が定数 κ よりも小さい成分は縮退するため、出力次元が制限され新規データを予測できない出力次元崩壊が起きる。

謝辞　本発表は、神谷之康氏、白川健氏、田中美里氏、青木俊太郎氏、間島慶氏、村木優介氏との共同研究に基づくものです。また、本研究は JSPS 科研費 JP20H05954、JP20H05705、JP21K17821、22KJ1801 及び JST CREST JPMJCR18A5、JPMJCR22P3、国立研究開発法人新エネルギー・産業技術総合開発機構（ＮＥＤＯ）の助成事業（JPNP20006）の助成を受けたものです。

参考文献

[1] Ken Shirakawa, Yoshihiro Nagano, Misato Tanaka, Shuntaro C Aoki, Kei Majima, Yusuke Muraki, and Yukiyasu Kamitani. Spurious reconstruction from brain activity. *arXiv*, 2024.

[2] Guohua Shen, Tomoyasu Horikawa, Kei Majima, and Yukiyasu Kamitani. Deep image reconstruction from human brain activity. *PLoS Comput. Biol.*, 15(1):1006633, 2019.

[3] Furkan Ozcelik and Rufin VanRullen. Natural scene reconstruction from fMRI signals using generative latent diffusion. *Sci. Rep.*, 13(1):15666, 2023.

[4] Yu Takagi and Shinji Nishimoto. High-resolution image reconstruction with latent diffusion models from human brain activity. In *CVPR*, pages 14453–14463, 2023.

[5] Abdulkadir Canatar, Blake Bordelon, and Cengiz Pehlevan. Out-of-distribution generalization in kernel regression. In *NeurIPS*, volume 34, pages 12600–12612, 2021.

クープマン作用素に基づく縮約動力学表現の学習と予測・制御への応用

河原 吉伸 [1,2]
[1] 大阪大学大学院情報科学研究科，[2] 理化学研究所革新知能統合研究センター
e-mail：kawahara@ist.osaka-u.ac.jp

複雑な動的現象のデータ駆動による解析や予測・制御において，非線形力学系のクープマン作用素を用いた表現に基づくアプローチが近年注目されている [1, 2, 3]．クープマン作用素は，広い範囲の非線形力学系に対して定義可能であることが知られており，その線形性などの性質から，データを用いた解析方法との相性も良く，データ駆動による科学研究の主要なアプローチの一つとしても盛んに議論されている．特に，流体分野で提案されたデータ解析手法である動的モード分解 [4, 5, 6] や，その一般化である拡張動的モード分解 [7] との関連が指摘されるようになり，その重要性は更に高まっていると言える．最近では，機械学習分野においても，種々の問題においてクープマン作用素による力学系表現を原理的に応用した研究も増えつつある [8, 9, 10, 11]．

クープマン作用素を用いた力学系表現において重要な概念として，クープマン作用素のスペクトルによる分解があげられる．つまり，一般に線形作用素が固有値・固有関数により展開可能である性質を利用して，力学系により表現される動力学を，固有値で決まる周期性・減衰率を持つ “モード” の重ね合わせとして表現する (クープマン・モード分解とも呼ばれる)．この表現は部分的には物理的にも厳密に意味付けされ，例えば，一部の固有関数による写像が非線形振動子の位相振幅縮約を与えることが知られている [12, 13, 14]．また，系の大域的な安定性などの議論を，クープマン作用素の固有関数を用いて汎用的に記述可能であることも知られている [15]．そのように，クープマン作用素のスペクトルを介した低次元化は，物理的に重要な動力学特性を保存した縮約モデルを獲得する汎用的なアプローチとしても重要である．

一方，拡張動的モード分解によるクープマン作用素の推定における原理的課題の一つとして，作用素が定義される関数空間をどのように与えるのか，という問題があげられる．というのも，考察対象となる力学系を順方向的に解析する場合とは異なり，データを起点としてその背後にある力学系の解析を行う場合，解析したい動力学を抽出するための作用素を定義できる関数空間は自明ではない．また一般に，そのような空間は，クープマン作用素に対して閉じた形で与える必要がある (クープマン不変空間)．その解決方法として，機械学習の方法論を適用したアプローチが議論されており，拡張動的モード分解において再生カーネルを用いた方法 [16] や，また著者らによる研究を契機としたニューラルネットワークによりクープマン不変空間を同時推定する方法 [17, 18] などが提案されている．

また，機械学習を用いた将来予測や制御 (強化学習) において，上記のような特性を原理的に利用することも有用であり近年注目されている．例えば，例えば著者らにより，縮約を介して安定性を組み入れた時系列データの長期予測 [19] や，動力学の対称性を利用したデータ拡張に基づくオフライン強化学習 [20]，クープマン作用素のスペクトルを利用した自然な動作生成を実現するオンライン強化学習 [21] などが提案されている．

謝辞　ここで紹介する研究の一部は，JST CREST「作用素論的データ解析に基づく複雑ダイナミクス計算基盤の創出」(JPMJCR1913) の支援により実施されたものである．

参考文献

[1] B.O. Koopman, "Hamiltonian systems and transformation in Hilbert space," *Proceedings of the National Academy of Sciences of the United States of America*, 17(5): 315–318, 1931.

[2] I. Mezić, "Spectral Properties of Dynamical Systems, Model Reduction and Decompositions," *Nonlinear Dynamics*, 41: 309–325, 2005.

[3] S.L. Brunton, M. Budisić, E. Kaiser, and J.N. Kutz, "Modern Koopman Theory for Dynamical Systems," *SIAM Review*, 64(2): 229–340, 2022.

[4] C.W. Rowley, I. Mezić, S. Bagheri, P. Schlatter, and D.S. Henningson, "Spectral analysis of nonlinear flows," *Journal of Fluid Mechanics*, 641: 115–127, 2009.

[5] P.J. Schmid, "Dynamic mode decomposition of numerical and experimental data," *Journal of Fluid Mechanics*, 656: 5–28, 2010.

[6] J.H. Tu, C.W. Rowley, D.M. Luchtenburg, S.L. Brunton, and J.N. Kutz, "On Dynamic Mode Decomposition: Theory and Applications," *J. Computational Dynamics*, 1(2): 391–421, 2014.

[7] M.O. Williams, I.G. Kevrekidis, and C.W. Rowley, "A Data-Driven Approximation of the Koopman Operator: Extending Dynamic Mode Decomposition," *Journal of Nonlinear Science*, 25: 1307–1346, 2015.

[8] K. Fujii, Y. Inaba and Y. Kawahara, "Koopman spectral kernels for comparing complex dynamics: Application to multiagent sport plays," *Proc. of the 2017 European Conf. on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML-PKDD'17)*, pp.127–139, 2017.

[9] T. Konishi and Y. Kawahara, "Stable Invariant Models with Koopman Spectra," *Neural Networks*, 165: 393–405, 2023.

[10] Y. Hashimoto, S. Sonoda, I. Ishikawa, A. Nitanda, and T. Suzuki, "Koopman-based generalization bound: New aspect for full-rank weights," *Proceedings of the 25th International Conference on Learning Representations (ICLR'24)*, 2024.

[11] R. Hataya and Y. Kawahara, "Glocal Hypergradient Estimation with Koopman Operator," arXiv:2402.02741, 2024.

[12] A. Mauroy, I. Mezíc, and J. Moehlis, "Isostables, isochrons, and Koopman spectrum for the action-angle representation of stable fixed point dynamics," *Physica D*, 261: 19-30, 2013.

[13] S. Shirasaka, W. Kurebayashi, and H. Nakao, "Phase-amplitude reduction of transient dynamics far from attractors for limit-cycling systems," *Chaos*, 27: 023119, 2017.

[14] A. Mauroy and I. Mezić, "Global computation of phase-amplitude reduction for limit-cycle dynamics," *Chaos*, 28: 073108, 2018.

[15] A. Mauroy and I. Mezić, "Global Stability Analysis Using the Eigenfunctions of the Koopman Operator," *IEEE Transactions on Automatic Control*, 61(11): 3356–3369, 2016.

[16] M.O. Williams, C.W. Rowley, and I.G. Kevrekidis, "A kernel-based method for data-driven koopman spectral analysis," *Journal of Computational Dynamics*, 2(2): 247-265, 2015.

[17] N. Takeishi, Y. Kawahara and T. Yairi, "Learning Koopman invariant subspaces for dynamic mode decomposition," *Advances in Neural Information Processing Systems*, 30: 1130–1140, 2017.

[18] T. Iwata and Y. Kawahara, "Neural dynamic mode decomposition for end-to-end modeling of nonlinear dynamics," *Journal of Computational Dynamics*, 10(2): 268–280, 2023.

[19] N. Takeishi and Y. Kawahara, "Learning Dynamics Models with Stable Invariant Sets," *Proceedings of the 35th AAAI Conf. on Artificial Intelligence (AAAI'21)*, pp.9782–9790, 2021.

[20] M. Weissenbacher, S. Sinha, A. Garg and Y. Kawahara, "Koopman Q-learning: Offline Reinforcement Learning via Symmetries of Dynamics," *Proceedings of the 39th International Conference on Machine Learning (ICML'22)*, PMLR 162: 23645–23667, 2022.

[21] M. Ohnishi, I. Ishikawa, K. Lowrey, M. Ikeda, S. Kakade and Y. Kawahara, "Koopman Spectrum Nonlinear Regulators and Efficient Online Learning," *Transactions on Machine Learning Research*, ISSN: 2835-8856, 2024.