

条件付き生成のための拡張フローマッチング

磯部 伸¹¹ 東京大学大学院数理科学研究科

e-mail: nobo0409@g.ecc.u-tokyo.ac.jp

1 導入—条件付き生成モデルにおける課題—

昨今社会全体で注目されている生成モデル，特に条件付き生成モデルは，入力，または条件を表すある領域 $\Omega \subset \mathbb{R}^k$ から，あるデータ空間 $D \subset \mathbb{R}^d$ 上の条件付き確率分布がなす空間 $\mathcal{P}(D)$ への写像と解釈できる．例えば，マテリアルズインフォマティクスの分野においては，HOMO-LUMO ギャップなどの物性値 $c \in \Omega$ を条件として，その物性値を持つ分子構造からなる条件付き確率分布 $\mu_c \in \mathcal{P}(D)$ (のサンプル $x \in D$ による近似) を出力する生成モデルの構築を目指している．一度このようなモデルが得られてしまえば，特定の物性値 c を持つ分子構造 x の候補を，生成モデルが出力する分布からのサンプルとして簡単に得ることができる．

ここで問題となるのは，この条件付き生成モデルの構成方法である．現在最も有力視されているのは，フローマッチング [1] と呼ばれる枠組みに基づいた，深層学習による近似アルゴリズムである．そこでは，観測可能な条件 $c \in \Omega$ 毎に，近似したい条件付き分布 μ_c へ向かう D 上のベクトル場 v を学習する (図 1)．典型的には，動的最適輸送計画問題を近似的に解くことで，ベクトル場に対する教師データを作成し，回帰問題として学習を定式化する．このアルゴリズムは，画像生成や音声合成の分野を中心に盛んに応用されている．

しかし，このアルゴリズムを分子構造生成に適用すると，観測できていない条件 c において，あまり精度の高い近似ができない．その理由の一つは，フローマッチングが生成するサンプル x_c が， c を連続的に変化させたときに大きく変化し得るからである (図 2 の水色で塗られた分布)．

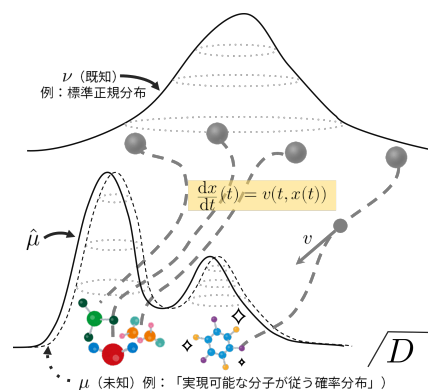


図 1: フローマッチングの概念図

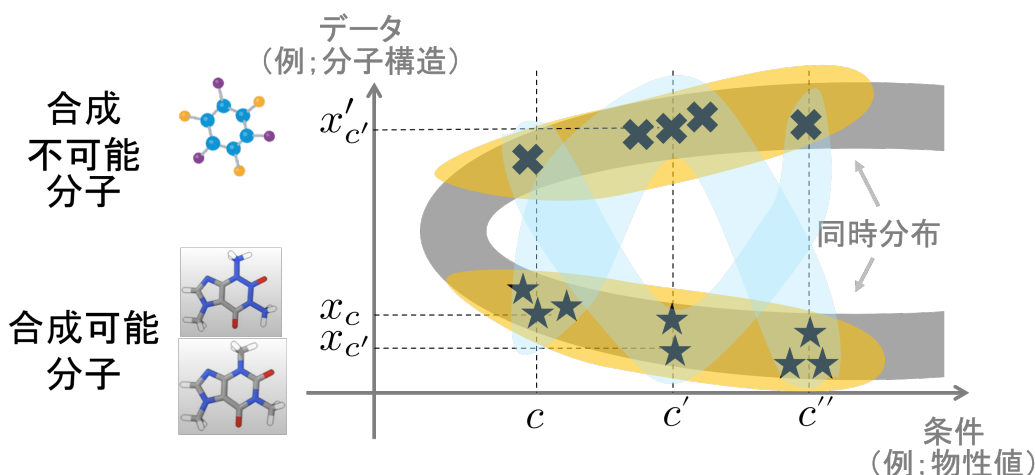


図 2: フローマッチングと拡張フローマッチングの生成結果の概念図. [2, Figure 1] も参照のこと.

2 条件付きフローマッチング

前述した問題を解決するために、我々は次の一般化連続方程式に対するある種の逆問題を解く：近似したい条件付き確率分布 $\mu: \Omega \ni c \mapsto \mu(\bullet | c) \in \mathcal{P}(D)$ から、学習データ $\left\{ (c_i, \frac{1}{N_x} \sum_{j=1}^{N_x} \delta_{x_{i,j}}) \right\}_{i=1}^{N_c} \subset \Omega \times \mathcal{P}(D)$ が得られているとする．このデータに対して、条件付き確率分布 $\hat{\mu}$ と行列値関数（行列場） $u: \Omega \times D \rightarrow \mathbb{R}^{d \times k}$ であって、 $\hat{\mu}(c_i) \approx \mu(c_i)$, $i = 1, \dots, N_c$ であり、方程式系

$$\nabla_c \mu(c) + \operatorname{div}_x(\mu(x | c) u(c, x)) = 0 \quad (c, x) \in \Omega \times D, \quad (2.1)$$

を満たすものの中から、(図 2 の橙色の分布のように)「あまり大きく変化しないもの」を見出せ．ここで、方程式系 (2.1) は超関数の意味で考え、勾配 ∇ や発散 div は、行列場と整合的になるように定義される．方程式系 (2.1) は、質量保存則をあらわす連続方程式、または連続の式における時間を、条件 c という多次元ベクトルに拡張したものになっている．

我々はこの問題に対して、以下に概略するアルゴリズムを提案した：

- (i) ((経験) 条件付き分布間のマッチング) : $\tilde{\mu}(c_i) = \frac{1}{N_x} \sum_{j=1}^{N_x} \delta_{x_{i,j}}$ となる $\tilde{\mu}: \Omega \rightarrow \mathcal{P}(D)$ であって、「あまり大きく変化しないもの」、ここでは、動的最適輸送計画問題の一般化である**確率測度値関数に対するディリクレ問題** [3] の近似解を、 $\tilde{\mu}(c) = \frac{1}{N_x} \sum_{i=1}^{N_x} \delta_{\psi^i(c)}$ の形で構成する．ここで $\psi^i: \Omega \rightarrow D$, $i = 1, \dots, N_x$ は「条件付きサンプル間のマッチング」を表す関数である．
- (ii) (行列場の学習) 方程式系 $\nabla_c \tilde{\mu}(c) + \operatorname{div}_x(\tilde{\mu}(x | c) \tilde{u}(c, x)) = 0$ を満たす行列場 \tilde{u} を $L(\theta) := \iint_{\Omega \times D} |u_\theta - \tilde{u}|^2$ の最小化により求める．ここで、 $u_\theta: \Omega \times D \rightarrow \mathbb{R}^{d \times k}$ はニューラルネットワークであり、 $\theta \in \mathbb{R}^p$ はそのパラメータである．

ステップ (ii) は一見不可能に見えるが、以下の定理により実装可能となる：

Theorem 2.1 ([2, Theorem 4] の概略)．目的関数 $L(\theta)$ は、

$$L'(\theta) := \int_{\Omega} \sum_{i=1}^{N_x} |u_\theta(c, \psi^i(c)) - \nabla \psi^i(c)|^2 \, dc$$

と、 θ に依存しない定数の違いを除いて一致する．

*1 本予稿は、小山雅典, Jinzhe Zhang, 林浩平 (Preferred Networks), 福水健次 (統計数理研究所) との共同研究に基づく．

参考文献

- [1] Y. Lipman, R. T. Q. Chen, H. Ben-Hamu, M. Nickel, and M. Le. “Flow Matching for Generative Modeling.” *The Eleventh International Conference on Learning Representations*. 2023. URL: <https://openreview.net/forum?id=PqvMRDCJT9t> (cit. on p. 1).
- [2] N. Isobe, M. Koyama, J. Zhang, K. Hayashi, and K. Fukumizu. “Extended Flow Matching: A Method of Conditional Generation with Generalized Continuity Equation.” 2024. arXiv: [2402.18839](https://arxiv.org/abs/2402.18839) [cs.LG]. URL: <https://arxiv.org/abs/2402.18839> (cit. on pp. 1, 2).
- [3] H. Lavenant. “Harmonic Mappings Valued in the Wasserstein Space.” *Journal of Functional Analysis* **277.3** 2019, pp. 688–785. URL: <https://www.sciencedirect.com/science/article/pii/S0022123619301478> (cit. on p. 2).

混合最適輸送理論とその応用

谷地村 敏明¹

¹ 東北大学数理科学共創社会センター

e-mail : toshiaki.yachimura.a4@tohoku.ac.jp

1 概要

最適輸送理論は確率測度間の距離や最適マッチングを提供する変分問題として広く知られている。近年、正則化最適輸送理論の進展により、Sinkhorn アルゴリズムと GPU を組み合わせた高速計算が可能となった。この技術の発展に伴って大規模データセットに対する最適輸送理論の適用が現実のものとなり、グラフィックス、自然言語処理、生命科学、機械学習など多岐にわたる分野でその応用が急速に拡大している [1]。一方で実際の問題に適用する際には、依然としてその計算コストの高さが課題となることが多い。こうした問題に取り組むために、ソース及びターゲットの確率測度が混合確率測度であると仮定する混合最適輸送理論が注目されている。混合最適輸送は異なるクラスターやパターンが存在するデータに対して、構造を考慮に入れた距離や最適マッチングを効率的に計算することが可能であり、色調変換やテクスチャ合成 [2]、量子化学における電子構造計算 [3]、生命科学における細胞分化の動態解析 [4] など様々な分野で応用されている。以下ではこの混合最適輸送問題が、あるベクトル値最適輸送問題と同値であることについて考察する。また、講演では混合最適輸送問題の正則化についても考察し、その細胞分化の動態解析 [4] への応用について簡単に紹介する。

2 混合最適輸送理論

\mathbb{R}^d 上の確率測度全体の集合を $\mathcal{P}(\mathbb{R}^d)$ とする。以下ではその部分集合 $\mathcal{A} \subset \mathcal{P}(\mathbb{R}^d)$ の混合確率測度と、混合確率測度全体の集合である $\mathcal{M}(\mathcal{A})$ を定義する。

定義 1 (混合確率測度). 確率測度の部分集合 $\mathcal{A} \subset \mathcal{P}(\mathbb{R}^d)$ について、 $\mu \in \mathcal{P}(\mathbb{R}^d)$ が \mathcal{A} の混合確率測度であるとは、 $K \in \mathbb{N}$, $a := (a^1, \dots, a^K) \in \mathcal{A}^K$, $\lambda := (\lambda^1, \dots, \lambda^K) \in \mathcal{L}_K$ が存在して、

$$\mu = \sum_{k=1}^K \lambda^k a^k$$

とかけることである。ここで $\mathcal{L}_K = \left\{ t = (t_1, \dots, t_K) \in \mathbb{R}^K \mid t_k \geq 0 \ (k = 1, \dots, K), \sum_{k=1}^K t_k = 1 \right\}$ である。 \mathcal{A} の有限混合確率測度全体の集合を $\mathcal{M}(\mathcal{A})$ とし、 \mathcal{A} -混合と呼ぶ。

\mathcal{A} の典型例として、非退化ガウス測度全体の集合 $\mathcal{A}_g^d := \{g_{m, \Sigma} \mid m \in \mathbb{R}^d, \Sigma \in \mathcal{S}^d\}$ が挙げられる。ここで \mathcal{S}^d は $\mathbb{R}^{d \times d}$ の正定値対称行列全体の集合である。このとき $\mathcal{M}(\mathcal{A}_g^d)$ は d 次元有限混合ガウス測度全体の集合を意味する。また、 (\mathcal{A}_g^d, W_2) は測地空間であることに注意する。本講演では、特に \mathcal{A}_g^d -混合 $\mathcal{M}(\mathcal{A}_g^d)$ について考察することにする。

定義 2 (混合 Wasserstein 距離 [3]). $\mu_0, \mu_1 \in \mathcal{M}(\mathcal{A})$ とする。このとき、

$$MW_2(\mu_0, \mu_1)^2 := \inf_{w \in \Pi(\lambda_0, \lambda_1)} \sum_{i=1}^{K_0} \sum_{j=1}^{K_1} w_{ij} W_2(\mu_0^i, \mu_1^j)^2$$

を $\mathcal{M}(\mathcal{A})$ 上の混合 Wasserstein 距離という．ここで $\Pi(\lambda_0, \lambda_1)$ は

$$\Pi(\lambda_0, \lambda_1) := \left\{ w := (w_{ij})_{1 \leq i \leq K_0, 1 \leq j \leq K_1} \in \mathbb{R}_+^{K_0 \times K_1} \left| \sum_{k=1}^{K_1} w_{ik} = \lambda_0^i (1 \leq i \leq K_0), \sum_{k=1}^{K_0} w_{kj} = \lambda_1^j (1 \leq j \leq K_1) \right. \right\}$$

である．実際，この混合 Wasserstein 距離 $MW_2(\mu_0, \mu_1)$ は $\mathcal{M}(\mathcal{A})$ 上の距離である．

3 混合 Wasserstein 距離とベクトル値最適輸送問題との関係

$M_n^+(\mathbb{R}^d)$ を \mathbb{R}^d 上の正の n 次元ベクトル値測度全体の集合とし， $\mathcal{P}_n(\mathbb{R}^d) = \{m = (m_1, \dots, m_n) \in M_n^+(\mathbb{R}^d) \mid \sum_{i=1}^n m_i \in \mathcal{P}(\mathbb{R}^d)\}$ とする．ここで， $\mu, \nu \in \mathcal{P}_n(\mathbb{R}^d)$ に対して，ベクトル値測度 $\gamma = (\gamma_{ij})_{1 \leq i, j \leq n} \in M_{n \times n}^+(\mathbb{R}^d \times \mathbb{R}^d)$ で $\mu_i(A) = \sum_{k=1}^n \gamma_{ik}(A \times \mathbb{R}^d)$ 並びに $\nu_j(B) = \sum_{k=1}^n \gamma_{kj}(\mathbb{R}^d \times B)$ ($A, B \subset \mathbb{R}^d$) を満たすものをベクトル値測度の輸送計画と呼び， μ と ν の輸送計画全体の集合を $\Pi(\mu, \nu)$ とする．

定義 3 (ベクトル値最適輸送問題 [5])．コスト関数 $c_{ij} : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ は正值連続関数とする．ベクトル値測度 $\gamma \in \Pi(\mu, \nu)$ に対する総輸送コスト $K(\gamma)$ として，

$$K(\gamma) = \sum_{i,j=1}^n \int \int_{\mathbb{R}^d \times \mathbb{R}^d} c_{ij}(x, y) d\gamma_{ij}(x, y)$$

と定める．このときベクトル値最適輸送問題とは， $K(\mu, \nu) := \inf_{\gamma \in \Pi(\mu, \nu)} K(\gamma)$ である．

混合 Wasserstein 距離とベクトル値最適輸送問題との関係として，以下が成り立つ：

定理 4. $\mu_0, \mu_1 \in \mathcal{M}(\mathcal{A})$ は $\mu_0 = \sum_{k=1}^n \lambda_0^k a_0^k =: \sum_{k=1}^n \mu_0^k$ ， $\mu_1 = \sum_{k=1}^n \lambda_1^k a_1^k =: \sum_{k=1}^n \mu_1^k$ であり， $c_{ij}(x, y) = |x - y|^2$ とする．このとき，

$$MW_2(\mu_0, \mu_1)^2 = K(\mu, \nu)$$

が成立する．

参考文献

- [1] G. Peyré, M. Cuturi, Computational optimal transport: With applications to data science, Foundations and Trends® in Machine Learning, Vol. 11, (2019), 355–607.
- [2] J. Delon, A. Desolneux, A Wasserstein-type distance in the space of Gaussian mixture models, SIAM Journal on Imaging Sciences, Vol. 13, (2020), 936–970.
- [3] G. Dusson, V. Ehrlacher, N. Nouaime, A Wasserstein-type metric for generic mixture models, including location-scatter and group invariant measures, arXiv:2301.07963, (2023).
- [4] T. Yachimura, H. Wang, Y. Imoto, M. Yoshida, S. Tasaki, Y. Kojima, Y. Yabuta, M. Saitou, Y. Hiraoka, scEGOT: Single-cell trajectory inference framework based on entropic Gaussian mixture optimal transport, bioRxiv, (2023).
- [5] X. Bacon, Multi-species optimal transportation, Springer, Vol. 184, (2020), 315–337.

正則化最適輸送問題におけるスケーリングの影響

榊原 航也¹, 高津 飛鳥², 保國 恵一³

¹ 金沢大学数物科学系, ² 東京都立大学理学研究科, ³ 筑波大学システム情報系
e-mail : ksakaki@se.kanazawa-u.ac.jp

1 Bregman ダイバージェンス正則化

最適輸送問題に対する数値計算手法は数多く提案されているが, エントロピー正則化を用いた効率的な数値計算アルゴリズムが近年発見されて以降, 機械学習など多くの分野で用いられている [1].

そのような中で, 我々の研究 [2] では, エントロピー正則化を一般化し, Bregman ダイバージェンスによる正則化を検討した. 以降, 記号の定義等は [2] を参照せよ. $U \in C([0, 1]) \cap C^1((0, 1))$ を $[0, 1]$ 上の狭義凸関数とすると, Bregman ダイバージェンス $D_U: \mathcal{P}_K \times \mathcal{P}_K \rightarrow \mathbb{R}$ は

$$D_U(z, w) := \sum_{k=1}^K (U(z_k) - U(w_k) - (z_k - w_k)U'(w_k))$$

そして, Bregman ダイバージェンス正則化

$$\inf_{\Pi \in \Pi(x, y)} (\langle C, \Pi \rangle + \varepsilon D_U(\Pi, x \otimes y)) \quad (\varepsilon > 0)$$

に対して, その最小化元を $\Pi^U(C, x, y, \varepsilon)$ と書くとき, 次の誤差評価が成り立つことを示した:

$$\langle C, \Pi^U(C, x, y, \varepsilon) \rangle - \inf_{\Pi \in \Pi(x, y)} \langle C, \Pi \rangle \leq \Delta_C(x, y) \cdot e_U \left(-\frac{\Delta_C(x, y)}{\varepsilon} + \mathfrak{D}_U(x, y) + \nu_U(x, y) \right). \quad (1.1)$$

ただし, e_U は U' の逆関数であり, また,

$$\begin{aligned} \mathfrak{D}_U(x, y) &:= \sup_{\Pi \in \Pi(x, y)} D_U(\Pi, x \otimes y), \\ \Delta_C(x, y) &:= \inf_{V' \in V(x, y) \setminus \arg \min_{V \in V(x, y)} \langle C, V \rangle} \langle C, V' \rangle - \inf_{V \in V(x, y)} \langle C, V \rangle, \\ \nu_U(x, y) &:= \sup_{r \in (0, R_U(x, y)]} (U'(1 - r) + rU''(r)) \end{aligned}$$

とおいた. ここで, $R_U(x, y) \in [1/2, 1)$ は $U'(R_U(x, y)) - U'(1 - R_U(x, y)) = \mathfrak{D}_U(x, y)$ を満たすものである. また, $U \in C([0, 1]) \cap C^1((0, 1)) \cap C^2((0, 1))$ であり, かつ $(0, 1)$ 上で $U'' > 0$, $\lim_{h \searrow 0} U'(h) = -\infty$, そして $(0, 1) \ni r \mapsto rU''(r)$ が非減少であることが仮定として必要である.

2 スケーリング

本講演では, データ x, y や U の定義域のスケーリングが誤差評価に与える影響を考察する.

2.1 データと関数を同時にスケーリングした場合

$b > 0$ とし, $W \in C([0, b]) \cap C^1((0, b))$ を $[0, b]$ 上の狭義凸関数とする. そして, $a > 0$ に対して, 関数 $W_b^a: [0, a] \rightarrow \mathbb{R}$ を $W_b^a(r) := ab^{-1}W(a^{-1}br)$ により定義し, 次の正則化問題を考える:

$$\inf_{\tilde{\Pi} \in \Pi(ax, ay)} \left(\langle C, \tilde{\Pi} \rangle + \varepsilon D_{W_b^a}(\tilde{\Pi}, ax \otimes y) \right).$$

$U := W_b^1$ とおくと,

$$\begin{aligned}\mathfrak{D}_{W_b^a}(ax, ay) &= a\mathfrak{D}_U(x, y), & R_{W_b^a}(ax, ay) &= aR_U(x, y), & \nu_{W_b^a}(ax, ay) &= \nu_U(x, y), \\ \Delta_C(ax, ay) &= a\Delta_C(x, y), & e_{W_b^a} &= ae_U\end{aligned}$$

が成り立つ. よって, (1.1) 式により,

$$\begin{aligned}\langle C, \Pi^{W_b^a}(C, ax, ay, \varepsilon) \rangle &- \inf_{\tilde{\Pi} \in \Pi(ax, ay)} \langle C, \tilde{\Pi} \rangle \\ &\leq \frac{\Delta_C(ax, ay)}{a} e_{W_b^a} \left(-\frac{\Delta_C(ax, ay)}{a\varepsilon} + \frac{1}{a} \mathfrak{D}_{W_b^a}(ax, ay) + \nu_{W_b^a}(ax, ay) \right)\end{aligned}$$

が成り立つ. すなわち, データ x, y と狭義凸関数の定義域を同時にスケールした場合, (1.1) 式に示している誤差評価と本質的に同じ評価を導くことができる.

2.2 データと関数を別々にスケールした場合

$\tilde{a} < a$ なる $\tilde{a}, a, b > 0$ をとり, 次の正則化問題を考える:

$$\inf_{\tilde{\Pi} \in \Pi(\tilde{a}x, \tilde{a}y)} \left(\langle C, \tilde{\Pi} \rangle + \varepsilon D_{W_b^a}(\tilde{\Pi}, \tilde{a}x \otimes y) \right).$$

まず,

$$\operatorname{argmin}_{\tilde{\Pi} \in \Pi(\tilde{a}x, \tilde{a}y)} \left(\langle C, \tilde{\Pi} \rangle + \varepsilon D_{W_b^a}(\tilde{\Pi}, \tilde{a}x \otimes y) \right) = \operatorname{argmin}_{\Pi \in \Pi(x, y)} \left(\langle C, \Pi \rangle + \varepsilon D_{W_b^{a/\tilde{a}}}(\Pi, x \otimes y) \right)$$

が成り立つので, データのスケールは狭義凸関数の定義域のスケールと等価である. ここで, Bregman ダイバージェンスについて, 次の命題が成り立つことがわかる.

命題 1. U は § 1 最後の条件を満たし, また $\tilde{a}, a > 0$ は $\tilde{a} < a \leq 1$ を満たすとする. もし

$$D_U(\tilde{a}z, \tilde{a}w) = \kappa D_U(az, aw) \quad \text{for } z, w \in \mathcal{P}_K \text{ with } K \geq 3$$

が成り立つような $\kappa > 0$ が存在するならば, ある $\mu_0, \mu_1 \in \mathbb{R}$, $\lambda > 0$ が存在し, $U(r) = \lambda r \log r + \mu_0 r + \mu_1$ ($r \in (0, a]$) が成り立つ.

この命題は, Bregman ダイバージェンスが Kullback–Leibler ダイバージェンスでない限り, スケールリングの仕方が誤差評価に影響を及ぼすことを示唆している.

謝辞 本研究は, JSPS 科研費 JP22K03425, JP22K18677, JP23H00086 (榊原), JP19K03494, JP19H01786 (高津), JP20K14356, JP21H03451 (保國) の助成を受けたものです.

参考文献

- [1] G. Peyré and M. Cuturi, Computational optimal transport: With applications to data science, vol. 5–6, pp. 355–607. Now Publishers, Delft, Netherlands (2019).
- [2] K. Morikuni, K. Sakakibara, and A. Takatsu, Error estimate for regularized optimal transport problems via Bregman divergence, arXiv:2309.11666, 2023.

包除積分ニューラルネットワークへの単調正則化項の導入

本田 あおい¹, 穴井 日菜¹, Simon James²

¹ 九州工業大学, ²Deakin University

e-mail: aoii@ai.kyutech.ac.jp

1 概要

包除積分ニューラルネットワークは、包除積分数理モデルをニューラルネットワークに組み込んだものであり、出力層の重みパラメータが積分に用いられる測度のメビウス別表現に対応している。このネットワークは、精度の高い予測を行うだけでなく、学習後の重みを解析することによって予測の理由を説明可能にする特性を有している。しかし、入力層の重みパラメータの符号が適切でない場合、出力との単調関係が逆転し、解釈が困難になるという問題がある。本研究では、この問題を解決するために、包除積分ニューラルネットワークに対する単調正則化項を提案する。

提案する正則化項はコスト関数に追加され、出力層の重みを測度の単調性に基づいて制約する。これにより、モデルの解釈性が向上するだけでなく、ロバスト性も強化され、過学習を抑制する効果が期待できる。本研究では、回帰データセットを用いた実験を通じて、提案手法がモデルの単調性を保持しつつ、正則化を加えない場合とほぼ同等の精度を達成できることを示す。

2 包除積分ニューラルネットワーク

包除積分は、測度の非加法性を反映することが可能なファジィ積分の一種であり、ルベーク積分やショック積分を特別な場合として包含する。この積分は包除原理に基づいて定義されており、分割による上限をとるような計算を含まないため、コンピュータアルゴリズムとの親和性が高いという特長を持っている。以下、包除積分の定義を示す。全体集合を $X = \{1, 2, \dots, n\}$ とし、 v を 2^X 上のファジィ測度とする。さらに $\mathbf{x} = \{x_1, x_2, \dots, x_n\} \in [0, 1]^n$ を X 上の有界な非負関数、 \otimes を $[0, 1]$ 上の n 項演算とする。

定義 1 (包除積分) $\mathbf{x} = \{x_1, x_2, \dots, x_n\} \in [0, 1]^n$ の μ 、 \otimes による包除積分は次で定義される。

$$\otimes \int \mathbf{x} dv := \sum_{A \in 2^X} \left(\sum_{B \supset A} (-1)^{|B \setminus A|} \otimes_{i \in B} x_i \right) v(A).$$

この包除積分は、 v のメビウス変換 μ^v を用いて次のように別表現することができる：

$$\otimes \int \mathbf{x} dv = \sum_{A \in 2^X} \left(\otimes_{i \in A} x_i \right) m^v(A), \quad m^v(A) := \sum_{B \setminus A} (-1)^{|A \setminus B|} v(B)$$

包除積分ニューラルネットワークでは、このメビウス逆表現を利用する。

3 単調正則化項の提案

一般の入力ベクトルは \mathbb{R}^n の値をとるが、包除積分の被積分関数、つまり包除積分層への入力は $[0, 1]^n$ である必要がある。このため、前処理層ではシグモイド関数、 $h(x) = 1/(1 + e^{-(ax+b)})$ 、を活性化関数として用い、包除積分層への入力を $[0, 1]^n$ に正規化する。包除積分層が積分として意味を持つためには、入力と出力が正の単調関係を維持する必要がある。そのため、学習データの各説明変数と目的変数の相関関係を調べ、包除積分層への各入力 x_i が出力と正の単調関係を持つようにパラ

メータ a の初期値の符号を設定する。しかし、学習データに偏りがある場合、目的変数と説明変数の単独な値に基づく相関係数が実際の相関関係を反映しないことがある。このような場合、適切でない初期値が設定され、学習を進めても入力の変数間の関係の逆転が矯正されず、そのため推論は可能であるものの、包除積分層が積分の計算として機能していない、といった事態が生じる。この問題を解決するため、正則化項をコスト関数に加えることを検討する。

まず、ファジィ測度が単調性、 $A \subset B \implies v(A) \subset v(B)$ を満たすことの必要十分条件をメビウス変換を用いて整理すると、任意の $i \in X$ に対して以下が成り立つ:

$$\sum_{C \subset A} m^v \geq \sum_{D \subset B} m^\mu(D), \quad m^v(A) \geq - \sum_{B \subset A, B \ni i} m^v(B).$$

したがって、 $m^v(A)$ について、

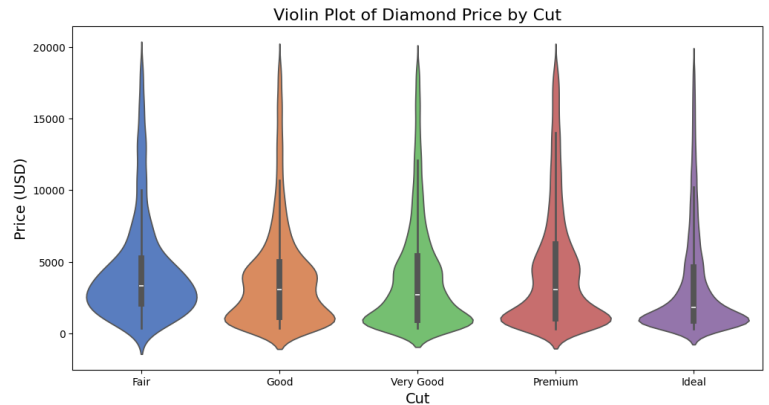
$$\Delta m^v(A) := m^v(A) + \min_i \left\{ \sum_{B \subset A, B \ni i} m^v(B) \right\} < 0$$

が成り立つときには単調性を満たしていないことになる。したがって、このような全ての $A \in 2^X$ に対して $\Delta m^v(A)$ が正になるように増加させる必要がある。これを正則化項で表現すると、 $w_A = m^v(A)$ の対応より、正則化項は次のように定義される:

$$R_{\text{MONO}} = \sum_{A \in 2^X, \Delta w(A) < 0} |\Delta w(A)|.$$

4 検証実験

単調正則化項の有効性を評価するために、Kaggle の Diamonds Prices データセットを用いて検証実験を行った。図は、このデータセットの price (価格、縦軸) と cut (カットの質、横軸) の関係を示したバイオリンプロットである。横軸はカットの質を表し、右に行くほどカットの質が高い。しかし、カットの質が単独で高い場合でも、必ずしも価格が高くなるわけではないことが示されている。



参考文献

- [1] Hina Anai, Simon James, Aoi Honda, Inclusion-exclusion integral neural network using monotone regularization terms, Proceedings of The 21th International Conference on Modeling Decisions for Artificial Intelligence, to appear.
- [2] Aoi Honda, Masayuki Itabashi, Simon James, A neural network based on the inclusion-exclusion integral and its application to data analysis, Information Sciences, Vol. 648, 119549, 2023.