具有无序目标的神经网络中的责任问题

Ben Hayes、Charalampos Saitis 与 Gy¨orgy Fazekas 伦敦
玛丽女王大学电子工程与计算机科学学院，英国{b.j.hayes、
c.saitis、george.fazekas}@qmul.ac.uk

arXiv:2304.09499v1 [cs.LG] 19 Apr 2023

摘要

我们讨论了将无序对象映射到具有固定排列的神经网络输出时产生的不连续性，这被称为责任问题。以往研究通过识别出单个不连续点，证明了该问题的存在性。在本文中，我们表明，在此类模型下，不连续点的数量是不可数无穷，从而推动对无序数据神经网络的进一步研究。

1 引言

责任问题（Zhang 等，2020b）描述了训练具有无序目标的神经网络时出现的一种问题：输出单元的固定排列要求每个单元为某个元素承担“责任”。责任排列会导致不连续性：在排列不变度量中很小的变化，会要求层的实际输出发生很大的变化。对于前馈网络而言，此类不连续函数的最坏情况逼近在输入空间的至少某个子集上会任意糟糕（Kratsios 和 Zamanlooy，2022）

在集合预测任务中，经验上已经观察到性能下降（Zhang 等，2020a），这促使研究者探索能够规避这些不连续性的集合生成架构（Zhang 等，2020a；Kosiorek 等，2020；Rezatofighi 等，2018）。文献中曾对该问题进行简要的形式化表述（Zhang 等，2020b），但仅通过单个不连续性的存在性加以证明。在此，我们证明责任问题的另一种表述，将无序到有序的映射分为与排序同构的操作和不同构于排序的操作。

第一种情况下，出现的间断是责任的“经典”排列——也就是说，最初由一个输出单元表示的元素，后来改由另一个输出单元表示。利用 Hemasinha & Weaver (2015) 的一项结果，我们发现，任何在元素之间保持一致排序的映射都是同构的，因此都会出现这种间断。第二种情况下，凡是映射未能实现一致排序之处，也会出现间断。由此我们证明，无论是排序映射还是非排序映射，其间断点集合都是不可数无穷。

这一结果与实验所揭示的证据一致：如果模型架构未明确考虑集合数据无序的结构，模型性能就会下降。这对任何本质上涉及集合预测的任务都有实际意义，包括目标检测、点云生成、分子图生成、语音分离等。

2 责任问题

令 Θ 表示由 $\mathbb{R}^d$ 中的元素构成、基数为 $|\boldsymbol{\theta}| = n$ 的集合之集合 $\boldsymbol{\theta}$，其中 $d \geq 2$:

$$\boldsymbol{\theta} \triangleq \left\{ \boldsymbol{x}_k \mid \boldsymbol{x}_k \in \mathbb{R}^d, k \in \{1, \ldots, n\} \right\} \in \Theta.$$

我们感兴趣的是映射集合 $\mathcal{F}$ $f\colon \Theta \to \mathbb{R}^{n \times d}$，其中所有 $f \in \mathcal{F}$ 都会将任意 $\boldsymbol{\theta} \in \Theta$ 中的每个元素分别分配给所得矩阵的每一行:

$$f(\boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{x}_{\pi(1)} & \boldsymbol{x}_{\pi(2)} & \ldots & \boldsymbol{x}_{\pi(n)} \end{bmatrix}^T, \qquad \forall f \in \mathcal{F}, \quad \forall \boldsymbol{\theta} \in \Theta, \quad \pi = \Pi_f(\boldsymbol{\theta}) \ ,$$

其中，$\Pi_f\colon \Theta \to S_n$ 描述了应用于 $\boldsymbol{\theta}$ 索引的排列，而 S_n 是阶为 n的对称群。这种映射表示将无序对象（例如集合）分配给神经网络层中的各个单元。因此，我们希望证明所有 $f \in \mathcal{F}$ 都必须是不连续的，首先给出一些定义。

定义 1. 若对于所有 $\boldsymbol{\theta} \in \Theta$，则称 $\Pi_f\colon \Theta \to S_n$ a sorting operationunder an order $\leq$:

$$\boldsymbol{x}_{\pi(1)} \leq \boldsymbol{x}_{\pi(2)} \leq \cdots \leq \boldsymbol{x}_{\pi(n)}, \quad \pi = \Pi_f(\boldsymbol{\theta}).$$

设 $\mathcal{V}$ 为一个有序向量空间，由 $\mathbb{R}^d$ 构成，并配备记作 $\ll$的全序。也就是说，$\mathcal{V} = (\mathbb{R}^d, \ll)$。

定义 2. 如果全序 $\ll$ 上的 $\mathcal{V}$ 满足以下条件，则称其为字典序:

$$\begin{array}{ll} 1. & \boldsymbol{x} = \boldsymbol{y} \implies \boldsymbol{x} \ll \boldsymbol{y} \qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad \forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{V} \\ 2. & \boldsymbol{x} \ll \boldsymbol{y} \wedge \boldsymbol{x} \neq \boldsymbol{y} \implies x_i < y_i \qquad i = \min\left\{j | x_j \neq y_j\right\}, \forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{V} \end{array}.$$

定义 3. $\mathcal{V}$的基集 $\{\boldsymbol{v}_i\}_{i\in\{1,\ldots,d\}}$ 称为非阿基米德的，如果 $0 \ll \boldsymbol{v}_d \ll \cdots \ll \boldsymbol{v}_1$且 $a\boldsymbol{v}_{i+1} \ll \boldsymbol{v}_i$, $\forall a \in \mathbb{R}$。

然后，我们回顾 Hemasinha & Weaver（2015）中的以下定理。

定理 1. 任意的全序 $\ll$ 在 $\mathbb{R}^d$ 上，其中 $d \geq 2$ 同构于非阿基米德基上的字典序排列。

因此，我们可以在不失一般性的前提下，继续假设 $\mathcal{V}$定义在采用字典序 $\ll$的非阿基米德基上。然后，我们可以证明以下引理（证明见附录 A）。

引理 1. 如果 Π_f 不是全序下的排序操作，那么 Π_f 根本不是排序操作。

因此，关于给定的 Π_f，以下陈述之一成立：(1) Π_f 同构于字典序下的排序操作；或 (2) Π_f 不是排序操作。

这引出了以下关于第二种情形下不连续性的引理（证明见附录 A.1）.

引理 2. 对于映射 $f \in \mathcal{F}$，如果 Π_f 不是排序操作，那么 f 是不连续的。

因此，我们现在只关注 Π_f 在 $\ll$ 下为排序操作的情形，从而得到以下引理（证明见附录 A.1）。

引理 3. 对于映射 $f \in \mathcal{F}$，如果 Π_f 是排序操作，那么 f 不连续。

通过穷尽法，责任问题的证明便显而易见。

定理 2. 所有 $f \in \mathcal{F}$ 都不连续。

证明。由引理 2 和引理 3 可得:

（f 是排序操作）$\vee$（f 不是排序操作）$\implies$ f 不连续 $\quad \forall f \in \mathcal{F}$

前件显然是一个重言式，因此所有 $f \in \mathcal{F}$ 都是不连续的，我们已经证明了责任问题的存在性。 □

此外，通过排序和非排序映射证明了定理 2，可得如下推论（证明见附录 A.1）。

推论 1. 对于映射 $f \in \mathcal{F}$，使得 f 不连续的 $\theta \in \Theta$ 集合是不可数无限集。

3 结论

我们提出了神经网络中责任问题的另一种表述，并首次证明了该问题存在于一个不可数无穷的不连续类中。这个问题的存在性及其范围促使我们进一步研究用于集合生成的神经网络架构，以及存在此类不连续性时常见优化算法的行为。

URM 声明

作者确认，本研究至少有一位主要作者符合 ICLR 2023 Tiny Papers Track 的 URM 标准。

参考文献

Rohan Hemasinha 和 James R. Weaver。有序向量空间。《线性代数及其应用》，468：171–183，2015 年 3 月。ISSN 00243795。doi：10.1016/j.laa.2014.05.028。URL https://linkinghub.elsevier.com/retrieve/pii/S002437951400319X。

Adam R. Kosiorek、Hyunjik Kim 和 Danilo J. Rezende。使用 Transformer 进行条件集合生成，2020 年 7 月。URL http://arxiv.org/abs/2006.16841。arXiv：2006.16841 [cs]。

Anastasis Kratsios 和 Behnoosh Zamanlooy。在分段连续函数中学习子模式。《Neurocomputing》，480：192–211，2022 年 4 月。ISSN 09252312。doi：10.1016/j.neucom.2022.01.036。URL https://linkinghub.elsevier.com/retrieve/pii/S092523122200056X。

S. Hamid Rezatofighi、Roman Kaskman、Farbod T. Motlagh、Qinfeng Shi、Daniel Cremers、Laura Leal-Taix´e 和 Ian Reid。《Deep Perm-Set Net：使用深度神经网络学习预测排列和基数未知的集合》，2018 年 10 月。网址：http://arxiv.org/abs/1805.00613。arXiv：1805.00613 [cs]。

Yan Zhang、Jonathon Hare 和 Adam Pr¨ugel-Bennett。《深度集合预测网络》，2020年4月a。网址 http://arxiv.org/abs/1906.06565。arXiv:1906.06565 [cs，stat]。

Yan Zhang、Jonathon Hare 和 Adam Pr¨ugel-Bennett。《FSPool：使用特征级排序池化学习集合表示》，2020 年 5 月 b。网址：http://arxiv.org/abs/1906.02795。arXiv：1906.02795 [cs、stat]。

A 证明

引理 1. 如果 Π_f 不是全序下的排序操作，那么 Π_f 根本不是排序操作。

证明。设 Π_f 是由 $\lhd$表示的某种序下的排序操作。那么 $\lhd$ 必然是一个全序，因为，

$$f(\boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{x}_{\pi(1)} & \boldsymbol{x}_{\pi(2)} & \dots & \boldsymbol{x}_{\pi(n)} \end{bmatrix} \iff \begin{array}{r} \boldsymbol{x}_{\pi(1)} \lhd \boldsymbol{x}_{\pi(2)} \lhd \cdots \lhd \boldsymbol{x}_{\pi(n)}, \\ \pi = \Pi_f(\boldsymbol{\theta}),\ \forall \boldsymbol{\theta} \in \Theta. \end{array}$$

也就是说，二元运算 $\lhd$ 必须对所有 $\boldsymbol{x} \in \mathbb{R}^d$ 都有定义，因此根据全性公理，它是一个全序。 □

A.1 主要证明

为了继续证明引理 2 和 3，我们定义 d_Θ: $\Theta \times \Theta \to \mathbb{R}$ 为如下形式的度量:

$$d_\Theta(\boldsymbol{\theta}, \boldsymbol{\theta}') = \left\| \sum_{\boldsymbol{x} \in \boldsymbol{\theta}} \phi(\boldsymbol{x}) - \sum_{\boldsymbol{x}' \in \boldsymbol{\theta}'} \phi(\boldsymbol{x}') \right\|_p ,$$

其中， ϕ: $\mathbb{R}^d \to \mathbb{R}^k$ 是一个从 $\boldsymbol{\theta}$ 中的元素映射到任意表示的连续映射。随后，(Θ, d_Θ)是一个度量空间。

引理 2. 对于映射 $f \in \mathcal{F}$，如果 Π_f 不是排序操作，则 f 是不连续的。

证明。我们希望证明，在 $\mathbb{R}^{n\times d}$上的弗罗贝尼乌斯范数 $\|\cdot\|_F$下，存在一个 $\epsilon>0$，使得对于所有 $\delta>0$，存在两个集合 $\boldsymbol{\theta}$、$\boldsymbol{\theta}'\in\Theta$，满足 $d_\Theta(\boldsymbol{\theta},\ \boldsymbol{\theta}')<\delta$且 $\|f(\boldsymbol{\theta})-f(\boldsymbol{\theta}')\|_F\geq\epsilon$。
假设对于某个 $f\in\mathcal{F}$，Π_f不是排序操作，那么 $\mathcal{V}$中必定至少存在一对元素，使得 Π_f违反定义 1。因此，存在 $\boldsymbol{a}$、$\boldsymbol{b}\in\mathcal{V}$，满足$\boldsymbol{a}\ll\boldsymbol{b}$和 $\boldsymbol{a}\neq\boldsymbol{b}$，使得存在如下定义的 $\boldsymbol{\theta}$、$\boldsymbol{\theta}'\in\Theta$：

$$\begin{aligned}\boldsymbol{\theta}&=\{\boldsymbol{a},\boldsymbol{b},\boldsymbol{u}_1,\boldsymbol{u}_2,\ldots,\boldsymbol{u}_{n-2}\}\\ \boldsymbol{\theta}'&=\{\boldsymbol{a},\boldsymbol{b},\boldsymbol{u}'_1,\boldsymbol{u}'_2,\ldots,\boldsymbol{u}'_{n-2}\}\end{aligned},$$

其中，由 f产生的排列具有以下形式：

$$\begin{aligned}f(\boldsymbol{\theta})&=[\ldots\quad\boldsymbol{a}\quad\ldots\quad\boldsymbol{b}\quad\ldots]^T\\ f(\boldsymbol{\theta}')&=[\ldots\quad\boldsymbol{b}\quad\ldots\quad\boldsymbol{a}\quad\ldots]^T,\end{aligned}$$

其中，$f(\boldsymbol{\theta})$中与 $\boldsymbol{a}$、$\boldsymbol{b}$对应的行索引，不一定等于 $f(\boldsymbol{\theta}')$中与 $\boldsymbol{a}$、$\boldsymbol{b}$对应的行索引。必须存在这样一对元素，因为如果对于所有包含 $\boldsymbol{a}$、$\boldsymbol{b}$的 $\boldsymbol{\theta}\in\Theta$，$\ll$都发生逆序，那么结果在重新命名元素后与 $\ll$ 同构。根据这种必要的顺序逆转，我们可以得到以下不等式，因为 $\boldsymbol{a}$和 $\boldsymbol{b}$中至多只有一个可能在两个矩阵中被分配到同一列：

$$\|f(\boldsymbol{\theta})-f(\boldsymbol{\theta}')\|_F^2\geq\min\left\{\min_{i,j}\|\boldsymbol{a}-\boldsymbol{u}_i\|_2^2+\left\|\boldsymbol{a}-\boldsymbol{u}'_j\right\|_2^2,\min_{i,j}\|\boldsymbol{b}-\boldsymbol{u}_i\|_2^2+\left\|\boldsymbol{b}-\boldsymbol{u}'_j\right\|_2^2\right\}.$$

将该量记为 ϵ^2。令 $A\subset\Theta$ 为包含 $\boldsymbol{a}$、$\boldsymbol{b}$ 的所有 $\boldsymbol{\theta}$ 构成的集合，其中定义 1 对于$\boldsymbol{a}$ 和 $\boldsymbol{b}$成立；令 $B\subset\Theta$ 为包含 $\boldsymbol{a}$、$\boldsymbol{b}$ 的所有 $\boldsymbol{\theta}'$ 构成的集合，其中定义 1 不成立。令$C=A\cup B$。则 C 在 d_Θ下构成一个度量空间，并且具有 A、$B\subset C$ 和 $A=B^c$。
因此，必然存在 A 中的一个序列 $(\boldsymbol{\theta}_m)_{m\in\mathbb{N}}$ 以及一个元素 $\boldsymbol{\theta}'\in\partial B$，使得 $\lim_{m\to\infty}\boldsymbol{\theta}_k=\boldsymbol{\theta}'$。
于是，我们可以找到 $\boldsymbol{\theta}\in A$、$\boldsymbol{\theta}'\in B$，使得对于所有 $\delta>0$，$d_\Theta(\boldsymbol{\theta}、\ \boldsymbol{\theta}')<\delta$。
定义了 ϵ后，我们就满足了不连续性判据。 □

为了证明引理 3，我们需要一个简单的中间结果。
引理 4。所有 $f\in\mathcal{F}$ 都是单射。

证明。假设存在一个映射 $g\in\mathcal{F}$，它不是单射。那么我们有：

$$\exists\boldsymbol{\theta},\boldsymbol{\theta}'\in\Theta:\qquad f(\boldsymbol{\theta})=f(\boldsymbol{\theta}')\wedge\boldsymbol{\theta}\neq\boldsymbol{\theta}'.$$

然而，根据定义，矩阵 $\{f(\boldsymbol{\theta})_i\mid i\in\{1,\ldots,n\}\}$ 的行集合等价于集合 $\boldsymbol{\theta}$，因此：

$$\boldsymbol{\theta}=\{f(\boldsymbol{\theta})_i\mid i\in\{1,\ldots,n\}\}=\{f(\boldsymbol{\theta}')_i\mid i\in\{1,\ldots,n\}\}=\boldsymbol{\theta}'\neq\boldsymbol{\theta},$$

这就产生了矛盾。 □

引理 3. 对于映射 $f\in\mathcal{F}$，如果 Π_f 是一种排序操作，那么 f 是不连续的。

证明。我们再次希望证明，在 Frobenius 范数 $\|\cdot\|_F$ 下，关于 $\mathbb{R}^{n\times d}$，存在一个$\epsilon>0$，使得对于所有 $\delta>0$，都存在两个集合 $\boldsymbol{\theta}$、$\boldsymbol{\theta}'\in\Theta$，满足 $d_\Theta(\boldsymbol{\theta},\boldsymbol{\theta}')<\delta$ 且$\|f(\boldsymbol{\theta})-f(\boldsymbol{\theta}')\|_F\geq\epsilon$。

考虑两个向量 $\boldsymbol{a}$、$\boldsymbol{b}\in\mathcal{V}$，使得对于 $\epsilon>0$，有 $\boldsymbol{b}=\boldsymbol{a}+\epsilon\boldsymbol{v}_j$，并且对某个 $j>1$，其中 $\{\boldsymbol{v}_i\}_{i\in\{1,\ldots,d\}}$ 是 $\mathcal{V}$的非阿基米德基。根据字典序的定义，我们有 $\boldsymbol{a}\ll\boldsymbol{b}$。因此，矩阵 $\boldsymbol{A}$ 在 f 作用下属于 Θ 的像：

$$\boldsymbol{A}=[\ldots\quad\boldsymbol{a}\quad\boldsymbol{b}\quad\ldots]^T\in f[\Theta].$$

根据 f 的单射性，存在唯一的 $\boldsymbol{\theta}\in\Theta$ 映射到 $\boldsymbol{A}$：

$$\boldsymbol{A}=f(\boldsymbol{\theta})\implies\boldsymbol{\theta}=\{\ldots,\boldsymbol{a},\boldsymbol{b},\ldots\}.$$

令 $B_{\delta'}(\boldsymbol{b})$ 为半径为 $\delta' > 0$、中心为 $\boldsymbol{b}$ 的开球。令 $N_{\boldsymbol{a}}, \{\boldsymbol{v} \mid \boldsymbol{v} \in \mathcal{V}, \boldsymbol{v} \ll \boldsymbol{a}, \boldsymbol{v} \neq \boldsymbol{a}\}$ 为 $\mathcal{V}$ 中所有在 $\ll$ 下严格小于 $\boldsymbol{a}$ 的元素构成的集合。我们希望找到一个元素 $\boldsymbol{c} \in B_{\delta'}(\boldsymbol{b}) \cap N_{\boldsymbol{a}}$。换言之，我们要寻找一个位于 $\boldsymbol{b}$ 周围的 δ'-球内，且在字典序 $\ll$ 下严格小于 $\boldsymbol{a}$ 的元素。取 $\boldsymbol{c} = \boldsymbol{b} - \tau \boldsymbol{v}_1$ 作为 $0 < \tau < \delta'$，即可找到这样的元素。

那么，另一个矩阵 $\boldsymbol{A}' = f(\boldsymbol{\theta}')$，其中 $\boldsymbol{\theta}' = \{\ldots, \boldsymbol{a}, \boldsymbol{c}, \ldots\}$，假设所有未指定的元素均与 $\boldsymbol{\theta}$ 中的元素相同，必定具有如下形式：

$$\boldsymbol{A}' = [\ldots \quad \boldsymbol{c} \quad \boldsymbol{a} \quad \ldots]^T = f(\boldsymbol{\theta}').$$

于是我们得到：

$$\|f(\boldsymbol{\theta}) - f(\boldsymbol{\theta}')\|_F = \|\boldsymbol{A} - \boldsymbol{A}'\|_F = \sqrt{\|\boldsymbol{a} - \boldsymbol{c}\|_2^2 + \|\boldsymbol{b} - \boldsymbol{a}\|_2^2 + \ldots} \geq \|\boldsymbol{b} - \boldsymbol{a}\|_2 = \epsilon.$$

根据 d_Θ 的定义，我们有：

$$d_\Theta(\boldsymbol{\theta}, \boldsymbol{\theta}') = \|\phi(\boldsymbol{b}) - \phi(\boldsymbol{c})\|_p .$$

根据 ϕ 的连续性，我们总能选取一个 δ'，使得对于任意小的 δ，$d_\Theta(\boldsymbol{\theta}, \boldsymbol{\theta}') < \delta$。因此，我们找到了一个 $\boldsymbol{\theta} \in \Theta$ 类，其中 f 不连续。 □

推论 1. 对于映射 $f \in \mathcal{F}$，使 f 不连续的 $\theta \in \Theta$ 集合是不可数无限的。

证明。首先假设在 $\ll$ 下，Π_f 是一种排序操作。在引理 3 的证明中，我们任意选取了一个 $\boldsymbol{a} \in \mathbb{R}^d$ 的值。因此，对于任意选择的 $\boldsymbol{a} \in \mathbb{R}^d$，我们都能找到一个 $\boldsymbol{\theta}$，使得 f 在此处存在不连续性。令 D_f 表示 f 不连续的 $\boldsymbol{\theta}$ 的集合，则有 $|D_f| \geq |\mathbb{R}^d| > \aleph_0$，所以 D_f 必定不可数。

在 $\ll$ 下，若 Π_f 不是排序操作，我们可以将 Θ 划分为两个子集：Θ_{s} 包含所有根据 $\ll$ 进行排序的 $\boldsymbol{\theta}$，而 Θ_{u} 包含所有不根据 $\ll$ 进行排序的 $\boldsymbol{\theta}$。

然后我们按照引理 3 的证明中的构造来找到 $\boldsymbol{\theta}$、$\boldsymbol{\theta}'$。接着，要么（i）$\boldsymbol{\theta}$ 和 $\boldsymbol{\theta}'$ 都属于 Θ_{s}，在这种情况下，根据引理 3，我们在 $\boldsymbol{\theta}$ 处得到不连续性；要么（ii）$\boldsymbol{\theta}$ 或 $\boldsymbol{\theta}'$ 中至少有一个属于 Θ_{u}，在这种情况下，根据引理 2，我们得到不连续性。对于 $\boldsymbol{a} \in \mathbb{R}^d$ 的所有选择，这些情况必有一个成立，因此我们再次得到 $|D_f| \geq |\mathbb{R}^d| > \aleph_0$，所以对于所有 $f \in \mathcal{F}$，D_f 必须是不可数的。 □