动力系统站点上的解释分离与粘合

Paul WangLIP6，索邦大学，
巴黎，法国 paul.wang@lip6.fr

arXiv:2603.17141v1 [math.CT] 17 Mar 2026

我们构造一个 Grothendieck 位，其对象是在 o-极小结构中的可定义集上定义的 Mealy机，覆盖则是可定义开浸入的联合满射族。在此位点上，我们定义解释预层——即配备可解释接口、并由“判定器”参数化的系统。我们证明，行为预层（按可观测输出等价取商）是分离的：全局解释由其局部限制唯一确定。我们还表明，粘合一般会失败——局部一致的解释未必能在全局上组装起来；并针对受限接口预层中的无状态解释系统，给出层性质成立的充要拓扑条件，即判定器的纤维稳健不连通。

1 引言

机器学习中的可解释性旨在解释模型为何 会产生特定输出。当前方法的一个显著特点是其固有的**局部性**：解释针对特定数据点、特定电路或特定特征方向给出，绝不针对整个模型。根本问题在于：当存在多个局部解释时，它们能否彼此协调，形成连贯的全局图景？

这正是层论所要探讨的问题。层为每个开集分配数据，并规定交叠部分上的数据必须如何一致；层条件体现了这样一条原则：相容的局部数据能够唯一地粘合为全局数据。若局部数据无法粘合，原则上可以通过层上同调来度量这种阻碍（§7）。

此前已有研究采用层论方法研究动力系统：Schultz、Spivak 和 Vasilakopoulou [22] 将系统建模为时间区间上的层；Goguen [9] 采用观测上的层；Joyal、Nielsen 和 Winskel [14] 采用路径上的预层。在这些研究中，格罗滕迪克拓扑被赋予一个“基”范畴，系统则被建模为层。另辟蹊径的是，Abramsky 和 Brandenburger [1] 利用层刻画量子基础中的非局域性与情境性：局部测量数据无法扩展为全局截面，这一阻碍可由层上同调度量。我们的粘合失败（反例 5.2）在结构上与之类似——相容的局部解释无法扩展到全局——但数学内容截然不同：他们的预层定义在测量情景上，而我们的预层定义在动力系统范畴上。在可解释性方面，Geiger等人 [8] 将因果抽象 发展为机制可解释性的框架；他们的忠实性概念描述单个抽象与模型之间的关系，而我们的框架研究多个 局部抽象（截面）何时能够复合。Bassan、Amir 和 Katz [3]从计算复杂性视角研究局部与全局可解释性。

我们采取了一种不同的方法：在一个范畴上定义格罗滕迪克拓扑，该范畴的对象是系统，其态射为开浸入。此类位点上的层条件将相对于系统表示的局部到整体原理形式化。据我们所知，此前尚未研究过这样的位点结构。

草稿 2026年3月19日

© P. Wang 本作品依据知识共享署名-非商业性使用-禁止演绎许可协议授权。

贡献。我们在 o-极小结构中的可定义集上开展工作（这一框架提供了温和的几何以及一类规范的开浸入）：

1. 我们证明，粘合基底上的齐次 Mealy 机范畴继承粘合性（定理 2.8），并构造一个由可定义开浸入构成的 Grothendieck 位（$\mathrm{Sys}_{\mathrm{oi}}$，Cov）（§2）。2. 我们定义了解释预层的层级，其参数由一个“判定器”给出，该判定器将系统的接口转换为可解释概念（§3）。3. 我们证明，行为预层 $\mathscr{F}^{\mathrm{beh}}_{j}$是分离的（§4）：全局解释由其局部限制所确定。4. 我们表明，粘合通常会失败（§5），并针对限制接口的无状态解释预层 $\mathscr{F}^{\mathrm{ri}}_{j}$. 1 5.3，给出其满足层性质的充要拓扑条件，该条件以判定器各纤维的稳健不连通性表示（定理 5.5）。5. 我们引入 ε-近似截面（§6），并利用海利定理（定理 6.3）表明，输出空间的维数控制高阶协调障碍的深度。

记号。 我们用 $\mathscr{S}=(S_b,\ S_a,\ I,\ O\ \alpha)$ 表示一个异构Mealy机，其前状态为 S_b，后状态为 S_a，输入为 I，输出为 O，动力学为 $\alpha\colon S_b\times I\to S_a\times O$。所有系统的 I和O均非空。

2 系统、位点与 o-极小情形

我们建立范畴框架：o-极小结构中可定义集上的Mealy机、作为态射的可定义开浸入，以及由此得到的Grothendieck 位点。

2.1 异构Mealy机

定义 2.1. 设$\mathscr{C}$为一个具有有限积的范畴。异构Mealy机是一个五元组$\mathscr{S}=(\mathrm{S}_b, \mathrm{S}_a, \mathrm{I}, \mathrm{O}\ \alpha)$，其 $\alpha\colon \mathrm{S}_b\times\mathrm{I}\to \mathrm{S}_a\times\mathrm{O}$，其中 S_b（前状态）与 S_a（后状态）可以不同；当限制到一组状态时，若其后状态可能离开该子集，便会自然出现这种情形（关于时变状态空间的范畴处理，参见 [23]）。当 $\mathrm{S}_b=\mathrm{S}_a=:\mathrm{S}$ 时，我们就得到齐次Mealy机 $(\mathrm{S}, \mathrm{I}, \mathrm{O}\ \alpha)$；在 Set 中，这些是自函子 $\mathrm{H}(\mathrm{X})=(\mathrm{X}\times\mathrm{O})^{I}$ 的余代数。

一个**态射** $\mathscr{S}\to\mathscr{S}'$ 是一个四元组（$f_b,\ f_a,\ f_I,\ f_O$），使得动力学交换图成立：

$$\begin{array}{ccc} S_b\times I & \xrightarrow{\ \alpha\ } & S_a\times O \\ {\scriptstyle f_b\times f_I}\downarrow & & \downarrow{\scriptstyle f_a\times f_O} \\ S'_b\times I' & \xrightarrow[\ \alpha'\ ]{} & S'_a\times O' \end{array}$$

我们将所得范畴记为 Sys，将固定（I,O）的齐次系统所构成的非满子范畴记为 $\mathrm{Sys}_{\mathrm{ho}}(I,O)$；在该子范畴中，状态对象之间的态射在前状态和后状态下相同。

2.2 o-极小结构

ACT 对动力系统的研究通常以 **Set**、拓扑空间或光滑流形范畴为基础。我们则以 Def(M) 为基础，它是 o-极小结构 M 中可定义集所构成的范畴；也可以使用其宽子范畴 Def_c(M)，其中态射为连续可定义映射（注 2.4）。其动机源于格洛腾迪克的一项经典观点：他在《纲领草图》 [10] 中提出，拓扑空间范畴对于几何而言过于狂野，并呼吁建立一套温和拓扑（驯性拓扑）理论：

> “拓扑空间” *[...]* 这一概念显然并不恰当 *[...]*。拓扑结构 *[...]* 使“温和”现象 *[...]* 与“野性”现象 *[...]*共存。正是后者占据主导地位，使得现今的“拓扑流形”概念 *[...]*并不恰当。此外，只要我们还没有一种能够让我们免受其害的“温和性原理”，就很难想象我们能对任意维数的“温和空间”形成怎样合理的认识。——格洛腾迪克，《纲领草图》 [10], §5

o-极小结构由范登德里斯 [6] 在 Pillay 和 Steinhorn 的工作基础上引入，恰好提供了这样一种“温和性原理”：关于一维可定义集的单一公理，便能确保所有维度上的温和性。

定义 2.2 ([6, 第 1] 章)。 一个扩张了 $(\mathbb{R}, <, +, \cdot)$ 的**o-极小结构** M，由一列 $(\mathscr{D}_n)_{n\geq 1}$ 的子集布尔代数构成；这些代数在乘积和投影运算下封闭，并且 $\mathscr{D}_n$ 包含所有半代数集，而 $\mathscr{D}_1$ 恰由区间和点的有限并构成。 $\mathscr{D}_n$ 中的元素称为**可定义集**；它们之间的函数称为**可定义函数**，当且仅当其图是可定义的。

$\mathscr{D}_1$ 上的条件体现了“适度原则”：它排除了康托尔集、填充曲线以及其他病态对象。就我们的目的而言，其关键推论如下：

1. 细胞分解。 每个可定义集$X \subseteq \mathbb{R}^n$ 都可以划分为有限个细胞 （具有简单形式的可定义集，且与开球典范同胚），并且通过细化划分，可以使X 上的每个可定义函数分段连续 [6, 第 3] 章。特别地，可定义开子集都是连通可定义开集的有限并。

2. 粘合性。 范畴 Def(M) 是黏着的 [15]：它具有拉回和沿单态射的推出，且这些满足范坎彭条件。证明利用了忠实嵌入 Def(M) $\hookrightarrow$ **Set**：拉回和沿单态射的推出在底层集合上计算，并且仍然可定义（可定义映射的像是可定义的），因此范坎彭条件归结为**Set**已知的粘合性。粘合性确保沿开浸入的推出保持单态射——这一条件将在余芽粘合中使用（定理 5.1）。

ˇ 3. 有限上同调。 Edmundo 和 Peatfield [7]发展了 o-极小结构上可定义层的切赫上同调理论，并采用有限可容许覆盖和有限单纯神经。我们的 $\mathrm{Sys}_{\mathrm{oi}}$并非他们意义上的可定义位点（它是一个系统范畴，而非单个可定义空间），但固定系统的覆盖具有有限的可定义神经，因此他们的方法为第 7 节讨论的上同调障碍提供了一个自然的起点。

每个可定义集合$X \subseteq \mathbb{R}^n$ 都继承自 $\mathbb{R}^n$ 的拓扑。**可定义开浸入** $\iota : U \hookrightarrow X$ 是一个可定义单射，其像是一个可定义开子集。

例 2.3. o-极小结构的标准例子：$\mathbb{R}_{\text{alg}}$（半代数集合——仅涉及多项式），$\mathbb{R}_{\text{an}}$（全局次解析集合——收敛幂级数），$\mathbb{R}_{\text{an, exp}}$（添加指数函数）。神经网络激活函数（ReLU、sigmoid、softmax）在 $\mathbb{R}_{\text{an, exp}}$ 中都是可定义的，因此自然适用于这一框架。事实上，o-极小性已直接应用于深度学习理论：Ji 和 Telgarsky [12] 运用 o-极小分析的工具证明了深度齐次网络上梯度流的方向收敛；Bolte 和 Pauwels [5]则利用可定义集的惠特尼分层，证明了随机梯度方法在函数值上的收敛性。

注 2.4。 若在 **Set** 上工作，则所有单射都可以作为态射，且不受拓扑控制：S_b 的任意划分都会给出一个覆盖，状态空间的几何性质也无关紧要。在 Def(M) 中，o-极小拓扑提供了规范的一类“局部”包含映射、具有有限个连通分支的有限覆盖，以及定理 5.5 的层刻画所需的温和性。代价是 Def(M) 不是笛卡尔闭的。

Def(M) 中的态射不必连续；我们用 Def_c(M) 表示由连续可定义映射组成的宽子范畴。位点构造既适用于 Def(M)，也适用于 Def_c(M)，因为开浸入本质上是连续的（即带有子空间拓扑的开子集的包含映射）。黏着性、分离和余芽粘合只用到 Def(M)。层刻画（定理 5.5）则需要 Def_c(M)：其证明用到了输入空间的紧致性和正规性，而这些性质取决于映射是否连续。

2.3 位点Sys_{oi}

现在我们将系统结构与 o-极小几何结合起来。

定义 2.5（系统的可定义开浸入）。 态射 $\iota\colon \mathscr{S}' \hookrightarrow \mathscr{S}$ in Sys（在 Def(M) 上）称为**可定义开浸入**，如果每个分量 $\iota_b, \iota_a, \iota_I, \iota_O$ 都是可定义开浸入。将 Sys_{oi} （“开浸入”）记作以这些态射为态射的宽子范畴。

定义 2.6（覆盖）。 由可定义开浸入构成的有限族 $\{m_\alpha\colon \mathscr{S}_\alpha \hookrightarrow \mathscr{S}\}$ 是**覆盖**，如果诱导映射联合满射：$\bigsqcup_\alpha (S_{\alpha,b} \times I_\alpha) \twoheadrightarrow S_b \times I$ 且 $\bigsqcup_\alpha (S_{\alpha,a} \times O_\alpha) \twoheadrightarrow S_a \times O$。

在$S_b \times I$上的联合满射性严格强于分别对因子满射（$\bigsqcup S_{\alpha,b} \twoheadrightarrow S_b$和$\bigsqcup I_\alpha \twoheadrightarrow I$分别满射）：乘积条件确保对于每个 $(s,i) \in S_b \times I$，都有某个单独的局部片同时覆盖s和i。

命题 2.7。 Sys_{oi} 存在拉回（按连通分支计算，即可定义开集的交），且上述覆盖定义了一个格罗滕迪克预拓扑 *[17, Ch. III]*，使得(Sys_{oi},Cov) 成为一个位点。

证明概要。 (T1) 同构平凡地构成覆盖。

(T2) 给定 $\mathscr{S}$ 的一个覆盖 $\{m_\alpha\}$ 和 $n\colon \mathscr{S}' \hookrightarrow \mathscr{S}$，其拉回的连通分支为 $(S_{\alpha,b} \cap S'_b) \times (I_\alpha \cap I')$；联合满射性从 $S_b \times I$限制到 $S'_b \times I'$。

(T3) 可定义开浸入的复合仍是可定义开浸入，且联合满射性具有传递性。

□

定理 2.8。 对于所有 I、$O \in$ Def(M)，范畴 $\text{Sys}_{\text{ho}}(I, O)$ 是黏着的。

这是一个更一般结果（附录 B）的特例：

定理 2.9。 设 $\mathscr{C}$ 为一个具有有限积的粘合范畴 *[15]*，且 $H : \mathscr{C} \to \mathscr{C}$ 是一个保持拉回的自函子。那么，余代数范畴 Coalg(H) 是粘合的。

对于 $\mathscr{C} =$ **Set**，Padberg [19] 已证明这一点；一般情形似乎是新的。粘合性确保粘合证明中的余极限构造会产生定义良好的齐次系统；结合严格始对象（空可定义集），还可得到可扩张性 [15, 引理 4.1]。

3 判定器与解释预层

对系统的解释 会将其动力学转化为人类可理解的表述。此处，我们将人类可解释性建模为相对于某个判定器而言：选择一个可解释接口，并提供一张从系统接口到该可解释接口的映射。

3.1 判定器与截面

定义 3.1（判定器）。对于 $\mathscr{S} = (S_b,\ S_a,\ I、O\ \alpha)$，其判定器 是可定义映射$j_I : I \to I'$、$j_O : O \to O'$组成的二元组$j = (j_I,\ j_O)$，其中$(I',\ O')$是**可解释接口**。

构造 3.2（解释预层）。固定一个判定器 $\mathrm{j} : (\mathrm{I},\mathrm{O}) \to (\mathrm{I}',\mathrm{O}')$。对于可定义开浸入 $\mathrm{m} : \mathscr{S}_U \hookrightarrow \mathscr{S}$，定义 $\tilde{\mathscr{F}}_j(\mathscr{S}_U)$ 为由二元组 $(\mathscr{S}', \psi)$ 构成的集合，其中 $\mathscr{S}' = (\mathrm{S}'_b, \mathrm{S}'_a, \mathrm{I}', \mathrm{O}', \alpha')$ 且ψ: $\mathscr{S}_U \to \mathscr{S}'$ 是一个接口连通分支为 $\psi_I = \mathrm{j}_I \circ \mathrm{m}_I$ 和 $\psi_O = \mathrm{j}_O \circ \mathrm{m}_O$ 的态射。动力学交换图

$$\begin{array}{ccc} S_{U,b} \times I_U & \xrightarrow{\alpha_U} & S_{U,a} \times O_U \\ {\scriptstyle \psi_b \times (j_I \circ m_I)} \downarrow & & \downarrow {\scriptstyle \psi_a \times (j_O \circ m_O)} \\ S'_b \times I' & \xrightarrow[\alpha']{} & S'_a \times O' \end{array}$$

表示 $\mathscr{S}'$ 通过判定器忠实地呈现了 $\mathscr{S}_U$ 的动力学。限制通过前合成进行：$(\mathscr{S}',\ \psi) \mapsto (\mathscr{S}',\ \psi \circ n)$，其中 n: $\mathscr{S}_{U'} \hookrightarrow \mathscr{S}_U$。

切片站点 $(\mathrm{Sys}/\mathscr{S})^{\mathrm{op}}$ oi 上的预层 $\tilde{\mathscr{F}}_j$ 是一个层（分离和粘合可归结为联合满射，因为截面是字面上的有序对 $(\mathscr{S}',\ \psi)$），但它过于精细：它会区分仅在无关紧要的内部结构上有所不同的解释。问题在于应取哪个商。

3.2 等价关系与预层层级

约定 3.3。在本文余下部分，所有**解释系统** $\mathscr{S}'$（截面的目标）均为齐次的（$S'_b = S'_a =: S'$），因此动力学可以迭代，且行为等价是良定义的。被解释的系统仍然是异质的，因为它们可以由彼此不一致的前状态和后状态的补丁给出。

定义 3.4（余芽等价）。若存在一个齐次系统 $\mathscr{S}''$，其接口为 $(\mathrm{I}',\mathrm{O}')$，存在单态射 $\mathrm{i}_k : \mathscr{S}'' \hookrightarrow \mathscr{S}'\mathrm{k}\ (\mathrm{k} = 1, 2)$，属于 $\mathrm{Sys}_{\mathrm{ho}}(\mathrm{I}', \mathrm{O}')$（即具有恒等接口分量，且 $\mathrm{i}_{k,b} = \mathrm{i}_{k,a}$），以及一个 φ: $\mathscr{S}_U \to \mathscr{S}''$（必然唯一），满足 $\mathrm{i}_k \circ \varphi = \psi_k$，则称 $\mathscr{S}_U$ 上的两个截面 $(\mathscr{S}'_1, \psi_1)$ 和 $(\mathscr{S}'_2, \psi_2)$ 余芽等价。记作 $(\mathscr{S}'_1, \psi_1) \sim (\mathscr{S}'_2, \psi_2)$。

这个名称体现了它与芽的类比：如果两个解释在目标侧共享一个共同的“本质核心”，它们便是一致的。这种结构与 Joyal–Nielsen–Winskel [14] 中双模拟的基于跨度的刻画相平行：两者都使用跨度作为等价见证，但前者使用单态射而非开映射，并且是在目标侧而非路径侧进行。

商 $\mathscr{F}_j :=$ $\tilde{\mathscr{F}}_j/\sim$ 满足粘合但不满足分离。

定义 3.5（行为等价）。在 $\mathscr{S}_U$ 上的两个解释 $(\mathscr{S}'_1, \psi_1)$ 和 $(\mathscr{S}'_2, \psi_2)$ 是**行为等价的**，记作 $\approx$；如果对每个 $x \in S_{U,b}$、每个 n 以及每个 $(i'_1, \ldots, i'_n) \in (I')^n$，$\psi_{1,b}(x)$ 在 $\mathscr{S}'_1$ 中和 $\psi_{2,b}(x)$ 在 $\mathscr{S}'_2$ 中产生的输出序列相同（这是良定义的，因为根据约定 3.3，$\mathscr{S}'_k$ 是齐次的）。

余芽等价蕴含行为等价，但反之不成立（例如，具有相同常值输出的 3-循环和 6-循环是行为等价的，但不是芽等价的）。对于 **Set** 上的齐次 Mealy 机，行为等价与标准的余代数概念相关，即在 $H(X) = (X \times O')^{I'}$ [20] 的终余代数中像相等。从可解释性的角度来看，行为等价是一种自然关系：解释通过判定器来预测行为，而作出相同预测的两个解释可以互换。

定义 3.6（受限接口预层）。对于$m\colon \mathscr{S}_U \hookrightarrow \mathscr{S}$，定义 $I'_U := \mathrm{im}(j_I \circ m_I) \subseteq I'$。受限接口**预层** $\mathscr{F}\mathrm{ri}_j(\mathscr{S}_U)$ 由解释 $(\mathscr{S}',\ \psi)$ 构成，其中 $\mathscr{S}'$ 具有输入接口 I'_U（而非整个 I'），并按受限行为等价取商（仅在 $(I'_U)^n$ 中测试输入序列）。

由限制输入接口给出的自然变换 $\eta\colon \mathscr{F}_j^{\mathrm{beh}} \to \mathscr{F}\mathrm{ri}_j$ 在截面上是满射的（$\mathscr{F}\mathrm{ri}_j$ 在 $\mathscr{S}_U$ 上的任意截面都可以通过在 $I' \setminus I'_U$ 上任意选择动力学，扩展为 $\mathscr{F}\mathrm{beh}_j$ 的一个截面；这不会影响受限行为等价类），并且当 $I'_U = I'$ 时是一个同构。实际上，这一区别对应着两种可解释性模式：$\mathscr{F}_j^{\mathrm{beh}}$ 的截面对应于机制性解释（“此特征能在所有输入中检测到 X”），而 $\mathscr{F}_j^{\mathrm{ri}}$ 的截面对应于局部解释（“此电路在分布 $\mathscr{D}$ 上实现 Y”）。

我们将 $\mathscr{F}\mathrm{beh}_j := \tilde{\mathscr{F}}_j/\!\approx$ 记作**行为预层**。其层级如下：

$$\tilde{\mathscr{F}}_j \twoheadrightarrow \mathscr{F}_j \twoheadrightarrow \mathscr{F}_j^{\mathrm{beh}} \xrightarrow{\eta} \mathscr{F}_j^{\mathrm{ri}}.$$

$\sim\mathscr{F}_j$ 是一个层，但过于精细；这三个商刻画了逐渐粗略的解释内容概念。它们都不是层（§4–5）。

4 分离

分离公理提出的问题是：如果两个全局截面在覆盖的每个成员上限制为同一个类，那么它们在全局上是否等价？

定理 4.1（行为预层是分离的）。 $\mathscr{F}\mathrm{beh}_j$ 在 $(\mathrm{Sys}_{\mathrm{oi}}, \mathrm{Cov})$ 上满足分离公理。

证明。 行为等价是逐点的：$(\mathscr{S}'_1, \psi_1) \approx (\mathscr{S}'_2, \psi_2)$ 当且仅当对于每个 $x \in S_{U,b}$，$\mathrm{Beh}_{\mathscr{S}'_1}(\psi_{1,b}(x)) = \mathrm{Beh}_{\mathscr{S}'_2}(\psi_{2,b}(x))$，其中 Beh 表示从状态到无限输出树的映射。如果这在联合满覆盖中的每个 $\mathscr{S}_\alpha$ 上都成立，那么对于每个 $x \in S_{U,b}$，都存在 α 使得 $x \in S_{\alpha,b}$，因此它们在 x 处的行为一致。 □

余芽预层 $\mathscr{F}_j$ 不满足分离性：局部余芽见证可能包含截面的像之外的“额外状态”，而这些额外状态在不同补丁之间可能不一致。转向行为商后，只需检验可观测输出序列，便可消除这一问题。

定义 4.2（j-满覆盖）：覆盖 $\{m_\alpha\colon \mathscr{S}_\alpha \hookrightarrow \mathscr{S}\}$ 称为 **j-满**，如果对于每个 α 都有 $I'_{U_\alpha} = I'_U$，即每个补丁的输入都命中 j_I 的每个纤维。满足 $m_{I,\ \alpha} = \mathrm{id}_I$ 的覆盖（我们称之为**数据局部**覆盖）自动是 j-满的。

命题 4.3（ri$j$ 的分离性）：$\mathscr{F}_j^{\mathrm{ri}}$ 对于任意 j-满覆盖都满足分离性。

证明。 当对于所有 α 都有 $I'_{U_\alpha} = I'_U$ 时，局部和全局的受限行为等价检验相同的输入序列 $(I'_U)^n$。在 S_b 上的联合满射性（由从 $S_b \times I$ 的投影得到）完成了论证。 □

当覆盖将输入分散到不同局部片时，分离可能失效：

反例 4.4。考虑 $\mathscr{S} = (\{s_1, s_2\}, \{s_1, s_2\}, \{a, b\}, \{0, 1\}, \alpha)$，其中 $\alpha(s_1, a) = (s_1, 0)$、$\alpha(s_1, b) = (s_2, 0)$、$\alpha(s_2, a) = (s_1, 1)$、$\alpha(s_2, b) = (s_2, 1)$，且 j 为 id。输入分割覆盖 $\mathscr{S}_1 = \mathscr{S}|_{\{a\}}, \mathscr{S}_2 = \mathscr{S}|_{\{b\}}$ 有 $I'_{U_1} = \{a\} \neq \{b\} = I'_{U_2}$。另一种解释 $\mathscr{S}'_2$ 采用了不同的内部路由，但在每个局部片的纯输入序列上与恒等解释一致，却在混合序列 (a,b) 上不一致。

5 粘合概览

虽然余芽预层 $\mathscr{F}_j$ 始终满足粘合公理（定理 5.1），但行为预层和受限接口预层并不满足。我们将梳理正反两方面的结果。

5.1 $\mathscr{F}_j$的粘合

定理 5.1. $\mathscr{F}_j$ 满足粘合公理：给定一个覆盖 $\{m_\alpha : \mathscr{S}_\alpha \hookrightarrow \mathscr{S}\}$ 以及一个相容截面族 $[\mathscr{S}'_\alpha, \psi_\alpha] \in \mathscr{F}j(\mathscr{S}_\alpha)$，则存在 $[\mathscr{S}', \psi] \in \mathscr{F}j(\mathscr{S})$，其限制到每个 $[\mathscr{S}'_\alpha, \psi_\alpha]$上均为相应截面。

证明概要。选择余芽见证 $\mathscr{S}''_{\alpha\beta}$，以确保交叠上的相容性。全局目标 $\mathscr{S}'$ 是 $\cdots \mathscr{S}'_{\alpha^i} \xleftarrow{\alpha} \mathscr{S}''_{\alpha\beta_i} \xrightarrow{\beta} \mathscr{S}'_\beta \cdots$ 在 Sys 中的余极限；它可由粘合范畴 Def(M) 中沿单态射的迭代推出构造得到。粘合性确保典范映射 $\mathscr{S}'_\alpha \hookrightarrow \mathscr{S}'$ 仍为单态射；随后，相容态射 ψ_α 便可通过联合满射性粘合为 $\psi : \mathscr{S} \to \mathscr{S}'$。 □

5.2 $\mathscr{F}_j^{\mathrm{beh}}$的粘合失败

反例 5.2（$\mathscr{F}$beh$_j$的粘合失败）。设 $\mathscr{S} = (\{s_0, s_1, s_2, s_3\}, \{s_0, s_1, s_2, s_3\}, \{a, b\}, \{0, 1\}, \alpha)$，其动力学为：

$$\begin{aligned} \alpha(s_0, a) &= (s_0, 0), & \alpha(s_0, b) &= (s_2, 0), \\ \alpha(s_1, a) &= (s_1, 1), & \alpha(s_1, b) &= (s_1, 1), \\ \alpha(s_2, a) &= (s_1, 1), & \alpha(s_2, b) &= (s_1, 1), \\ \alpha(s_3, a) &= (s_3, 1), & \alpha(\text{秒}_3, \text{字节}) &= (\text{秒}_2, 1), \end{aligned}$$

带有判定器 j_I：$\{a, b\} \to \{\bullet\}$、$j_O = \mathrm{id}$。覆盖 $\mathscr{S}_1 = \mathscr{S}|_{\{s_0, s_1, s_2\}}, \mathscr{S}_2 = \mathscr{S}|_{\{s_1, s_2, s_3\}}$ 存在非平凡交叠 $\{s_1, s_2\}$，并且是数据局部的。

在 $\mathscr{S}_1$ 上：最小实现 $M_1 = (\{p_0, p_1\}, \{\bullet\}, \{0, 1\})$ 将 $s_0 \mapsto p_0$（常值输出 0）和 $s_1, s_2 \mapsto p_1$（常值输出 1）映射到相应状态。在 $\mathscr{S}_2$ 上：实现 $M_2 = (\{q\}, \{\bullet\}, \{0, 1\})$ 将 $s_1, s_2, s_3 \mapsto q$（常值输出 1）映射到相应状态。在 $\{s_1, s_2\}$ 上满足相容性：两者都产生常值输出 1。

全局截面（M $\bar{\psi}$）将要求：$\bar{\psi}$（s_0）产生输出 0，而 $\bar{\psi}$（s_3）产生输出 1。在（s_0, b）处的动力学交换图迫使 $\bar{\psi}_a$（s_2）继续产生输出 0；在（s_3, b）处，它迫使 $\bar{\psi}_a$（s_2）继续产生输出 1。矛盾：单个后继状态不可能产生这两种行为。

阻碍在于，动力学会将来自不同行为区域的状态（不同局部片中的s_0 和s_3，）映射到同一个后继状态s_2。相容性约束交叠部分 $\{s_1, s_2\}$，却不约束从交叠之外 到达的后继状态。j_I 的非单射性至关重要：将输入折叠会迫使每个初始状态仅对应一个动力学方块约束。

5.3 正面结果：连通纤维与鲁棒不连通纤维

将无状态子预层记作 $\mathscr{F}\mathrm{ri}_{j,1} \subseteq \mathscr{F}\mathrm{ri}_j$：满足 $|S'| = 1$的截面（$\mathscr{S}'$，ψ）。（限制保持$S'| = 1$，因此这确实是一个子预层；该记号可推广为$\mathscr{F}^{\mathrm{ri}}_{j,\,N}$，其中 $|S'| \le N$。）无状态截面可化为一个函数h：$I_U \to O'$，且$h = g \circ j_I$，其中g唯一确定；于是，粘合转化为一个局部到整体问题，即连续可定义函数对 $\ker(j_I)$ 的不变性问题。我们在 o-极小环境中给出精确刻画。

定义 5.3 (鲁棒不连通纤维)。 若存在开集 N $I^{-1}t_0$，使得I'中的两个不交非空开集V、W满足j_I^{-1} $(N) \subseteq V \cup W$，且$V \cap$与$j_I^{-1}(t_0) \neq \varnothing$不交，$W \cap$与$j_I^{-1}(t_0) \neq \varnothing$不交，则称纤维$j_I^{-1}(t_0)$为**鲁棒不连通**。

连通纤维绝不会鲁棒不连通。反之则不成立：纤维可以不连通，却不鲁棒不连通，因为这种不连通性会在邻近纤维中“愈合”。

例 5.4 (去心正方形)。 设$I = [0,1]^2 \setminus \{$为$(0.5,0.5)\}$，且$j_I(x,y) = x$。在$x = 0.5$ 处的纤维不连通，但邻近纤维 $\{x\} \times [0,1]\,(x \neq 0.5)$是连通的，且不能被$V \cup W$分离。不存在鲁棒不连通，因此尽管纤维不连通，$\mathscr{F}^{\mathrm{ri}}_{j,}$1仍是一个层。

定理 5.5 （无状态解释的层刻画）。 在Def$_c$（M）（具有连续可定义映射的宽子范畴）中，设 I 是紧致的，且im（j_O）是可定义连通的，并且 |im（j_O）| ≥ 2。那么

$$\mathscr{F}^{\mathrm{ri}}_{j,1} \text{ is a sheaf} \iff \text{no fiber of } j_I \text{ is robustly disconnected.}$$

当im（j_O）全不连通时，层条件显然成立。

证明概要。（⇒）给定一个稳健不连通 $\mathrm{j}_I^{-1}(\mathrm{N}) \subseteq \mathrm{V} \cup \mathrm{W}$，利用 j_O的连通像和可定义隆起函数，构造一个连续映射 $\mathrm{f}: \mathrm{I} \to \mathrm{O}$，使其在 V 和 W 上分别为 $\ker(\mathrm{j}_I)$-不变，但满足 $\mathrm{j}_O(\mathrm{f}(\mathrm{a})) \neq \mathrm{j}_O(\mathrm{f}(\mathrm{b}))$，其中 $\mathrm{a} \in \mathrm{V} \cap \mathrm{j}_I^{-1}(\mathrm{t}_0)$，$\mathrm{b} \in \mathrm{W} \cap \mathrm{j}_I^{-1}(\mathrm{t}_0)$。覆盖 $\mathrm{U}_1 := \mathrm{V} \cup (\mathrm{I} \setminus \mathrm{j}_I^{-1}(\mathrm{N}))$、$\mathrm{U}_2 := \mathrm{W} \cup (\mathrm{I} \setminus \mathrm{j}_I^{-1}(\mathrm{N}))$ 给出相容的局部截面，但这些截面无法粘合。

(⇐) 若h 局部具有 $\ker(j_I)$-不变性，但$h(x) \neq h(y)$，且$j_I(x) = j_I(y) = t_0$, 将$h\big|_{j_I^{-1}(t_0)}$ 的水平集分隔在不相交开集中（紧致I的正规性），那么利用紧致性可找到$N \ni t_0$，使$j_I^{-1}(N)$ 包含于这些开集的并集——这就是稳健不连通。 □

推论 5.6. 若 $j_I \Rightarrow$ 的连通纤维不存在稳健不连通，则 $\Rightarrow \mathscr{F}^{\mathrm{ri}}_{j,\,1}$ 是一个层。第一个蕴含是严格的（例 *5.4*）。

5.4 粘合概览总结

预层	分离?	粘合?
$\tilde{\mathscr{F}}_j$（未取商）	Yes	是（层）
$\mathscr{F}_j$（余芽）	No	Yes
$\mathscr{F}^{\text{行为的}}_j$（行为的）	Yes	否（反例 5.2）
$\mathscr{F}^{\mathrm{ri}}_j$ (受限接口)	j完全限定	通常不行
$\mathscr{F}^{\mathrm{ri}}_{j,}$ 1（无状态）	j完全限定	定理 5.5：当且仅当不存在稳健不连通性。

规律是：更细的等价关系允许粘合，却会破坏分离；更粗的等价关系实现分离，却会破坏粘合。“正确”的商取决于可解释性任务。

6 近似截面与 Helly 阻碍

精确的层条件要求考察局部一致的解释能否粘合成一个全局解释。在实践中，解释很少是精确的：探针只能达到给定的准确率，LIME 拟合具有忠实度界，概念瓶颈会泄露信息。我们将动力学方格从精确相等放宽为 ε-近似，并提出问题：局部近似一致的 解释何时能够组合成一个全局近似一致的解释?

6.1 ε-截面

设 (O', ρ) 为度量空间。对于无状态系统 $\mathscr{S}_U$，即 $|S_{U,b}| = |S_{U,a}| = 1$ 的情形，记 $f : I_U \to O$ 为输出函数 $f(i) := \pi_O(\alpha_U(*, i))$，其中 π_O 是到 O 的投影。$\mathscr{S}_U$ 上的无状态解释截面 ($|S'| = 1$) 可化为满足动力学交换图 $g(j_I(i)) = j_O(f(i))$ 的函数 $g : I' \to O'$，对所有 $i \in I_U$ 均成立。

定义 6.1 (ε-截面)。在 $\mathscr{S}_U$ 上的一个 **ε-截面** 是满足以下条件的函数 $g : I' \to O'$

$$\rho(g(j_I(i)),\, j_O(f(i))) \le \varepsilon \qquad \text{for all } i \in I_U.$$

记 $\mathscr{F}^{\varepsilon}_j(\mathscr{S}_U)$ 为 ε-截面的集合。注意，g 定义在完整接口 I' 上，而非受限的 $I'_U := \mathrm{im}(j_I|_{I_U})$ 上；这是无状态行为预层的 ε-松弛（$\mathscr{F}^{\mathrm{beh}}_j$ 的无状态截面，其中 $|S'| = 1$），而不是 $\mathscr{F}^{\mathrm{ri}}_{j,1}$ 的松弛。在 $\varepsilon = 0$ 处，这恢复了 $\mathscr{F}\mathrm{beh}_j$ 的精确无状态截面。

对于每个可解释输入 $i' \in I'_U$，定义**目标集**

$$T(i', U) := \{ j_O(f(i)) : i \in I_U,\ j_I(i) = i' \} \subseteq O'$$

以及**可行集** $A_\varepsilon(i', U) := \bigcap_{t \in T(i',U)} \bar{B}(t\ \varepsilon)$，即以目标值为中心的闭合 ε-球的交集。存在一个 ε-截面，当且仅当对每个 $i' \in I'_U$，都有 $A_\varepsilon(i', U) \neq \varnothing$。

对于覆盖 $\{U_1, \ldots, U_n\}$，目标集按并集组合：$T(i', \bigcup_\alpha U_\alpha) = \bigcup_\alpha T(i', U_\alpha)$，因此全局可行集可分解为

$$A_\varepsilon\big(i', \textstyle\bigcup_\alpha U_\alpha\big) = \bigcap_\alpha A_\varepsilon(i', U_\alpha). \tag{1}$$

在 i' 处存在全局截面，因而要求各局部可行集的交集非空。

注 6.2. $\mathscr{F}^{\varepsilon}_j$ 是一个层：限制对函数而言是恒等映射（全局函数 $g : I' \to O'$ 在每个局部片上要么是 ε- 接近的，要么不是），因此分离性质显然成立；而粘合性质则源于 ε- 约束是逐点施加于 $i \in I_U$ 的。这一层性质依赖于无状态限制；多状态 ε-截面的预层结构会更复杂。这与精确行为预层 $\mathscr{F}^{\mathrm{beh}}_j$ 形成对比，后者是分离的，但不是层（§4–5）。

6.2 障碍深度

在 $\varepsilon = 0$，成对截面存在性意味着全局存在性：如果每对局部片都存在公共精确截面，那么目标集全都是单点（同一个点），因此全局截面存在。在 $\varepsilon > 0$，这一结论不成立：可行集 $A_\varepsilon(i', U_\alpha)$ 是一些球，即使完全交集为空，它们的成对交集也可能非空。这正是海利定理所描述的情形。

定理 6.3 (海利阻碍深度)。 假设 $O' \subseteq \mathbb{R}^d$ 是凸的， ρ 是欧几里得度量。设$\{U_1, \ldots, U_n\}$为一个覆盖，其中 $n \geq d+1$，并固定 $i' \in I'$。如果对于每个$(d+1)$元子覆盖，都有 $A\varepsilon(i', U_{\alpha_1} \cup \cdots \cup U_{\alpha_{d+1}}) \neq \varnothing$，那么 $A\varepsilon(i', U_1 \cup \cdots \cup U_n) \neq \varnothing$。

界 $d+1$ 是紧的（例 *6.4*）。

证明。每个 $A\varepsilon(i', U\alpha) \cap O'$都是凸集（它是闭球与 $\mathbb{R}^d$ 中一个凸集的交集）。根据 (1)，对于任意子集 S，都有 $A\varepsilon(i', \bigcup_{\alpha \in S} U\ \alpha) = \bigcap_{\alpha \in S} A\ \varepsilon(i', U\alpha)$。假设表明， $\{A\varepsilon(i', U\alpha) \cap O'\}_\alpha$的每个含 $(d+1)$ 个元素的子族都有非空交集。由 $\mathbb{R}^d$中的海利定理 [11]，整个族的交集非空。 □

例 6.4 （等边三角形）。 令 $O' = \mathbb{R}^2$、$I = \{a, b, c\}$、$j_I \equiv \star$（所有输入都坍缩）、$j_O = \mathrm{id}$，以及 $f(a) = (0,0)$、$f(b) = (2,0)$、$f(c) = (1, \sqrt{3})$（边长为 2 的等边三角形）。三个单点补丁$U_\alpha = \{\alpha\}$。

成对切比雪夫半径（线段的外接半径）为 1；三元切比雪夫半径（三角形的外接半径）为 $2/\sqrt{3} \approx 1.155$。对于 $\varepsilon \in (1,\ 2/\sqrt{3})$:

每一对补丁上都存在• ε-截面（成对切比雪夫半径为 $= 1 \leq \varepsilon$），

• 但三者整体上不存在 ε-截面（$2/\sqrt{3} > \varepsilon$）。

这是 $\mathbb{R}^2$ 中真正的三重阻碍，无法归结为任何成对失败。这一构造可以推广： $\mathbb{R}^d$ 中的正则（$d+1$）单纯形会产生（$d+1$）重阻碍。

6.3 可解释性阅读

Helly 阻碍有一种直接的可解释性解读：即使每一对局部解释都能协调为一个联合的 ε 近似解释，整个集合仍可能在全局上无法协调一致；而输出空间的维数决定了必须同时检查多少个解释。

对于**硬** k类分类（$O' = \{1, \ldots, k\}$，离散度量），Helly 数为 2：两两可协调总是足够的。对于**软** 分类（$O' = \Delta^{k-1}$， $\mathbb{R}^{k-1}$中的概率单纯形），Helly 数为k：必须同时检查所有k 个局部解释。对于 $k \geq 3$，软分类器的协调复杂度严格高于硬分类器。

对于不受限制的多状态解释，若 ψ_b 是单射（$|S'| \geq |S_{U,b}|$, $\psi_b = \mathrm{id}_{S_{U,b}}$），则平凡划分（每个状态各自构成一类）会使各状态间的输出约束解耦。障碍深度仍至多为 $\dim(O')+1$，每个状态分别计算，各状态的贡献彼此独立。因此，无状态情形（$|S'| = 1$）是**最坏情形**：所有状态共享一个解释状态，从而产生最大的约束耦合。深入理解这一点仍有待未来研究。

7 讨论

具体实例化。 将神经网络视为动力系统：状态空间 S_b 是给定层的隐藏激活空间，动力学 α 将激活–输入对映射到下一层激活和网络输出，而判定器 $j = (j_I, j_O)$则将原始输入/输出空间投影到可解释概念上。于是，激活空间某一区域上的截面便构成一种解释：一个更简单的系统 $\mathscr{S}'$，通过判定器借助可解释接口 (I', O')忠实地表示网络的动力学。

有几种可解释性方法体现了这一模式。*Representation Engineering* [25]识别出“读取向量”$v \in \mathbb{R}^d$，使得隐藏状态$h \in S_b$ 在h 上的投影v^T ·能够追踪一个高层

概念（诚实性、情感、事实性）。在我们的框架中，j_O 将网络输出投影到概念得分上，而无状态截面（$|S'|=1$）是一个函数$g : I' \to O'$，仅根据可解释输入预测概念得分——这正是一个探针。自动机提取 [24] 能够恢复一个有限状态机$\mathscr{S}'$，该机器在形式语言上模拟 Transformer：这是一个多状态截面，其中解释系统有 $|S'|>1$ 个内部状态，而判定器将词元投影到有限字母表。特征归因方法（显著图 [2], 、SHAP [16]）为输入特征分配重要性得分；Bilodeau 等人 [4] 表明，这类方法可能会被证明失效——用我们的语言来说，特征归因解释的预层在某些区域上可能没有截面。**判定器是核心设计选择**：它形式化了“可解释”的含义，而不同的判定器会产生具有不同局部到整体性质的不同预层。

结果的含义。 分离 (定理 4.1)指出，如果两个行为解释在每个数据切片上都一致，那么它们在全局上也完全相同。这是一个非平凡的预测：能够在局部区分两个解释的行为探针，也能在全局区分它们——不存在“聚合导致的不可见性”。粘合失败 (反例 5.2)表明，局部一致的解释未必能拼合成一个全局解释。其阻碍具体而言是：动力学将来自不同的行为区域的状态映射到同一个后继状态，从而在交汇点对解释提出过多约束。这与多义性的表现类似：当一个神经元响应多个互不相关的概念时，“每个神经元对应一个概念”的解释在每个概念的数据区域上都成立，却无法在全局上粘合起来。两者的机制不同——多义性源于激活空间中的线性代数叠加，而粘合障碍则源于动力学——但局部成功、全局失败的模式是相同的。

层刻画（定理 5.5）将问题转化为拓扑问题：在连续可定义情形下，$\mathscr{F}^{\mathrm{ri}}_{j,\cdot}\,1$ 是一个层，当且仅当 j_I 的每个纤维都不是稳健不连通的。具体而言，如果可解释输入投影将原始输入分组成簇，那么当这些簇拓扑连通时，粘合成立；若某个簇裂分为无法通过附近簇弥合的连通分支，粘合便会失败。

ε-松弛（§6）针对的是解释仅为近似值的实际情形。最坏情况误差不超过 ε （在 O' 上的度量）的探针是一个 ε-截面。（这是一致界，比实践中常见的平均情况准确率要求更严格；Helly 界适用于最坏情况。）定理 6.3 进而给出如下预测：对于输出为 d 维的回归探针（$d \geq 2$），仅检查局部探针的两两一致性并不足够，必须检查 $(d+1)$-重一致性。对于硬分类（O' 离散），两两检查就足够（目标值必须全部一致）；对于软 k 类分类器（$O' = \Delta^{k-1} \subset \mathbb{R}^{k-1}$，凸），定理 6.3 给出的 Helly 数为 k。由此得到一项具体且可检验的预测：局部可解释性审计何时会漏掉整体不一致性。

范畴框架的作用。上述结果是通过直接的拓扑与组合论论证得到的：分离是逐点的，粘合障碍是一个明确的反例，层刻画利用了紧致性与正规性，而 Helly 界则运用了凸几何。范畴框架——开浸入的位、预层层级、覆盖条件——并不提供证明技巧；它提供的是问题。若没有层论语言，就没有理由去问局部探针能否粘合，也无法系统地区分分离与粘合，更没有框架将 Helly 界与层条件联系起来。目前使用的理论工具超出了证明实际所需：粘合性仅用于余芽粘合（定理 5.1），与可解释性相关的结果并未用到它。我们认为，这是一个新方向的首篇论文所具有的特点：框架的价值在于梳理概念图景；我们期待未来的工作（多状态刻画、图局部性、上同调障碍）能让范畴工具发挥关键作用。

抽象框架。这些结果可以清晰地归结为一对对象（$\mathscr{C}$，$\mathscr{M}$）上的抽象公理，其中 $\mathscr{C}$ 是有限完备、广延且正则的范畴，而 $\mathscr{M} \subseteq \mathrm{Mono}(\mathscr{C})$ 是一类在拉回与复合下封闭的单态射（扮演“开浸入”的角色）。每个结果只用到其所需的公理：分离（定理 4.1）只要求覆盖族联合满射——每个状态—输入对都能被某个局部片观察到——以及行为等价的逐点性质；余芽粘合（定理 5.1）用到 $\mathscr{C}$ 的完全粘合性（附录 B）；该性质对 Def(M) 成立，但对 Def$_c$(M) 不成立（备注 A.5）。将余芽粘合推广到连续情形——例如，将余芽见证限制为 $\mathscr{M}$-态射——仍是一个开放问题；不过，与可解释性相关的结果（分离、层刻画）并不需要这一性质；$\mathscr{F}^{\mathrm{ri}}_{J,\ 1}$ 的层刻画（定理 5.5）还要求具有拓扑富集 T：$\mathscr{C} \to \mathbf{Top}$，它将 $\mathscr{M}$-态射映为开嵌入，并且 $T(I)$ 具有紧致性与正规性。正则性与下降条件——有限联合满射的 $\mathscr{M}$-族是有效满态的——使这些覆盖构成一个次正则的格罗滕迪克预拓扑（命题 A.3）。

具体范畴 Def(M) 满足除 (T) 外的所有公理：它是黏着的，具有严格初始对象（因此是广延的），是正则的（像分解是可定义的），并且 $\mathscr{M}$ =具有可定义开浸入。这提供了除层刻画之外的所有结果。子范畴 Def$_c$(M) 还满足 (T) 和 (A’)，因而可以得到层刻画；但完全粘合性并不成立（备注 A.5），所以在该情形下尚未确立余芽粘合；Def$_c$(M) 是否正则仍有待验证（标准像分解显然无法直接迁移，因为连续的可定义双射未必是同胚），不过本文没有任何结果依赖这一点。Def$_c$(M) 中的开覆盖推出就是可定义开子集的并，而 VK 性质可借助粘贴引理（命题 A.4）归结为 **Set** 的粘合性；这足以满足预拓扑公理，却不足以支持余芽粘合中的余极限构造。这种因子分解阐明了每项结果真正依赖的条件，并将该框架拓展到其他基范畴：有限 T_0 空间或 p 进可定义集。

˘上同调方向。当粘合失败时，切赫上同调可以测量这种阻碍。反例 5.2 中的˘相容族表明，行为预层可能存在非平凡的切赫 1-余循环，用以记录动力学在交叠处所迫使出现的不相容行为类。要使这一点更加精确，需要定义系数结构：行为预层是集合预层，而非阿贝尔群预层，因此需要谨慎处理非阿贝尔切赫上同调或采取线性化步骤。对于有界线性化预层（将目标状态空间截断为 $|S'| \le N$），我们预计，在反例 5.2 的覆盖上，对于 $N \ge 2$，$\check{H}^1 = 0$：非正式地说，限制映射是满射，因为总能容纳一个新的源状态，而集合的满射会诱导自由向量空间的满射。这样，粘合障碍便会转移到$\check{H}^0$,，用于测量那些彼此相容、却无法全局组装的局部数据；如何精确刻画这一点是未来研究的一个方向。

超越本文。我们构建的位点捕捉数据局部性（将状态、输入和输出空间分解）。可解释性局部性的另外两个维度——图局部性（分解计算图）和特征局部性（分解激活空间）——需要不同的位点结构：图局部性要求复合分解的位点（在算子代数系统理论 [21, 18] 的意义下），而特征局部性则要求用于判定器的粗化的商网站。开发这些结构并研究它们之间的相互作用，是未来工作的重要方向。

参考文献

[1] Samson Abramsky 与 Adam Brandenburger（2011）：非局域性与语境性的层论结构。《新物理学杂志》13(11)，第113036页，doi:10.1088/1367-2630/13/11/113036。[2] Julius Adebayo、Justin Gilmer、Michael Muelly、Ian Goodfellow、Moritz Hardt 与 Been Kim（2018）：显著图的合理性检验。载于：《神经信息处理系统进展》，31，Curran Associates, Inc.。可见 https://proceedings.neurips.cc/paper_files/paper/2018/hash/294a8ed24b1ad22ec2e7efea049b8737-Abstract.html。[3] Shahaf Bassan、Guy Amir 与 Guy Katz（2024）：局部与全局可解释性：计算复杂性视角。载于：第41届国际机器学习会议论文集（ICML）。[4]Blair Bilodeau、Natasha Jaques、Pang Wei Koh 与 Been Kim（2024）：特征归因的不可能性定理。《美国国家科学院院刊》121(2)，第e2304406120页，doi:10.1073/pnas.2304406120。[5] Jérôme Bolte 与 Edouard Pauwels（2021）：保守集值场、自动微分、随机梯度方法与深度学习。《数学规划》188，第19–51页，doi:10.1007/s10107-020- 01501-5。[6] Lou van den Dries（1998）：《驯性拓扑与o-极小结构》。伦敦数学会讲义系列248，剑桥大学出版社，doi:10.1017/CBO9780511525919。[7] Mário J. Edmundo 与 Nicholas J. Peatfield（2008）：o-极小切赫上同调。《数学季刊》59(2)，第213–220页，doi:10.1093/qmath/ham033。[8] Atticus Geiger、Duligur Ibeling、Amir Zur、Thomas Icard 与 Joseph Y. Halpern（2025）：因果抽象：机制可解释性的理论基础。《机器学习研究杂志》。即将发表。[9] Joseph A. Goguen（1992）：并发交互对象的层语义。《计算机科学中的数学结构》2(2)，第159–191页，doi:10.1017/S0960129500001420。[10] Alexander Grothendieck（1997）：《纲领草图》。载于 Leila Schneps 与 Pierre Lochak 编：《几何伽罗瓦作用》，伦敦数学会讲义系列242，剑桥大学出版社，第5–48页，doi:10.1017/CBO9780511758874.003。写于1984年。[11] Eduard Helly（1923）：关于具有公共点的凸体集合。《德国数学家联合会年报》32，第175–176页。[12] Ziwei Ji 与 Matus Telgarsky（2020）：深度学习中的方向收敛与对齐。载于：《神经信息处理系统进展》（神经信息处理系统会议），33，第17176–17186页。[13] Peter Johnstone、John Power、Toru Tsujishita、Hiroshi Watanabe 与 James Worrell（2001）：论余代数范畴的结构。《理论计算机科学》260(1–2)，第87–117页，doi:10.1016/S0304- 3975(00)00124-9。[14] André Joyal、Mogens Nielsen 与 Glynn Winskel（1996）：来自开映射的双模拟。《信息与计算》127(2)，第164–185页，doi:10.1006/inco.1996.0057。[15] Stephen Lack 与 Paweł Sobociński（2005）：黏着的与拟黏合范畴。《RAIRO：理论信息学与应用》39(3)，第511–545页，doi:10.1051/ita:2005028。[16] Scott M. Lundberg 与 Su-In Lee（2017）：解释模型预测的统一方法。载于：《神经信息处理系统进展》，30，Curran Associates, Inc.，第4765–4774页。[17] Saunders Mac Lane 与 Ieke Moerdijk（1994）：《几何与逻辑中的层：托扑斯理论初步》。Universitext，施普林格出版社，doi:10.1007/978-1-4612-0927-0。[18] Nelson Niu 与 David I. Spivak（2025）：多项式函子：交互的数学理论。伦敦数学会讲义系列498，剑桥大学出版社。[19]Julia Padberg（2017）：重访分层图变换。载于：图变换（第十届国际图变换会议），计算机科学讲义10373，Springer，第20–35页，doi:10.1007/978-3-319-61470-0_2。扩展版本：arXiv:1702.04650。

[20] J. J. M. M. Rutten (2000)：泛余代数：一种系统理论。《理论计算机科学》249(1)，第 3–80 页，doi:10.1016/S0304-3975(00)00056-6。[21] Patrick Schultz 和 David I. Spivak (2019)：时态类型论：一种面向系统与行为的托扑理论方法。doi:10.1007/978-3-030-00704-1。[22] Patrick Schultz、David I. Spivak 和 Christina Vasilakopoulou (2020)：动力系统与层。《应用范畴结构》28，第 1–57 页，doi:10.1007/s10485-019-09565-x。[23] David Sprunger 和 Shin-ya Katsumata (2019)：通过延迟迹实现可微因果计算。载于：第 *34* 届 *ACM/IEEE* 计算机科学逻辑年度研讨会（*LICS*）论文集，第 1–12 页，doi:10.1109/LICS.2019.8785670。[24]Yihao Zhang、Zeming Wei 和 Meng Sun (2024)：从变换器中提取自动机(arXiv:2406.05564)。doi:10.48550/arXiv.2406.05564。可查阅：http://arxiv.org/abs/2406.05564。[25] Andy Zou、Long Phan、Sarah Chen、James Campbell、Phillip Guo、Richard Ren、Alexander Pan、Xuwang Yin、Mantas Mazeika、Ann-Kathrin Dombrowski、Shashwat Goel、Nathaniel Li、Michael J. Byun、Zi- fan Wang、Alex Mallen、Steven Basart、Sanmi Koyejo、Dawn Song、Matt Fredrikson、J. Zico Kolter 和 Dan Hendrycks (2023)：表征工程：实现人工智能透明性的自顶向下方法。doi:10.48550/ARXIV.2310.01405。可查阅：https://arxiv.org/abs/2310.01405。

A 抽象框架

我们为 §7 中描述的公理化框架给出精确的表述和证明。

定义 A.1（可容许对）。 一个可容许对 $(\mathscr{C},\ \mathscr{M})$由以下部分组成：

(C1) 一个有限完备、广延（有限余积互不相交且具有普遍性）、正则（每个态射都可分解为正则满态射 $+$ 单态射，且正则满态射在拉回下稳定）的范畴 $\mathscr{C}$；

(C2) 一个类 $\mathscr{M} \subseteq \mathrm{Mono}(\mathscr{C})$，在拉回和复合下稳定，并包含所有同构。

对象 B 的一个 $\mathscr{M}$-覆盖族 是一个有限族 $\{m_i\colon A_i \hookrightarrow B\}_{i=1}^{n}$，其中各态射均为 $\mathscr{M}$-单态射，且联合满射（也就是说，任意两个态射 $g,\ h\colon B \to C$ 若在每个 A_i 上相等，则必相等；在 Def(M) 等具体范畴中，这等价于它们联合满射）。

定义 A.2（模组化附加条件）。 给定一个可容许对 $(\mathscr{C},\mathscr{M})$：

(C3') 下降： 每个 $\mathscr{M}$-覆盖族都是有效满态的（即 B 是该覆盖族的 Čech 神经的余极限）。**(A')** 开覆盖范坎彭： 对于任意有限族 $\mathscr{M}$-子对象 $\{m_i\colon\ A_i \hookrightarrow X\}$，其中它们属于同一个对象 X，沿两两交叠 $A_i \times_X A_j$ 的推出 $\bigcup_i A_i$ 在 $\mathscr{C}$ 中存在，并满足范坎彭条件（沿任意态射拉回都会得到一个推出）。**(T)** 拓扑富集： 一个忠实函子 $T\colon\ \mathscr{C} \to \mathbf{Top}$，将 $\mathscr{M}$-态射映至开嵌入。

命题 A.3（子正则预拓扑）。 若$(\mathscr{C},\mathscr{M})$ 满足***(C1)–(C2)*** 和***(C3')***，则 $\mathscr{M}$-覆盖族在 $\mathscr{C}$ 上构成一个子正则格罗滕迪克预拓扑。

证明。 我们验证预拓扑公理。

（*T1*）同构构成覆盖。 $\{\mathrm{id}_B\}$ 是一个 $\mathscr{M}$-族（根据（C2），同构 $\in \mathscr{M}$），并且显然是联合满态射的。

(T2) 在拉回下的稳定性。 设 $\{m_i : A_i \hookrightarrow B\}$ 是一个覆盖，且$f :C \to B$ 是一个态射。拉回投影p_i：$C \times_B A_i \to C$是 $\mathscr{M}$-单态射（$m_i \in \mathscr{M}$的拉回，根据（C2）具有稳定性）。它

仍需证明 $\{p_i\}$是联合满射。根据（C3'），$\{m_i\}$ 是有效满态；根据可扩张性，这等价于典范映射$c:\bigsqcup_i A_i \to B$ 是一个有效满态（该族的 Čech 神经与c的核对相同）。特别地，c 是正则满态射，因此根据正则性在拉回下稳定：f^*c 是正则满态射。再次根据可扩张性，$f^*c \cong \bigsqcup_i (C \times_B A_i) \to C$ 是 $\{p_i\}$的典范映射，而正则范畴中的正则满态射是有效的，因此 $\{p_i\}$ 是联合满射。

(T3) 传递性。$\mathscr{M}$ 在复合下封闭 (C2)；联合满射族的复合仍是联合满射（若对所有i，j，有$g \circ m_i \circ n_{ij} = h \circ m_i \circ n_{ij}$，则由 $\{n_{ij}\}$ 的联合满射性可得$g \circ m_i = h \circ m_i$，进而由 $\{m_i\}$ 的联合满射性可得$g = h$）。

子典范性。有效满态族按定义满足下降：覆盖上的相容数据可以唯一地粘合。

□

命题 A.4 (Def 的开覆盖范坎彭$_c$(M))**.** 设M是一个 o-极小扩张$(\mathbb{R}, <, +, \cdot)$，$\mathrm{Def}_c(\mathrm{M})$ 是由可定义集合及连续可定义映射构成的范畴，$\mathscr{M} = \{$可定义开浸入$\}$。则$(\mathrm{Def}_c(\mathrm{M})\ \mathscr{M})$ 满足***(A')***。

证明。设U_1, U_2是可定义集合X的可定义开子集，且$V = U_1 \cap U_2$。我们证明$P = U_1 \cup U_2$（赋予从X继承的子空间拓扑）是$V \hookrightarrow U_1$和$V \hookrightarrow U_2$在 $\mathrm{Def}_c(\mathrm{M})$ 中的推出。

泛性质。给定连续可定义的 $g_1 : U_1 \to Z$ 和 $g_2 : U_2 \to Z$，且满足 $g_1|_V = g_2|_V$，定义 $g : P \to Z$，使得对于 $x \in U_i$，有 $g(x) = g_i(x)$。此定义在 V 上良定，并由粘贴引理保证连续（U_1 和 U_2 在 P 中是开集），且是可定义的：$\mathrm{graph}(g) = \mathrm{graph}(g_1) \cup \mathrm{graph}(g_2)$ 是有限个可定义集的并。

范坎彭。遗忘函子 $U : \mathrm{Def}_c(M) \to \mathrm{Set}$ 保持有限极限。集合论推出 $U(P) = U(U_1) \cup U(U_2)$ 与 Set 中的推出一致，而该推出因 Set 的粘合性而满足范坎彭条件。对于任意 $f : W \to P$，拉回给出可定义开子集 $W_i = f^{-1}(U_i) \subseteq W$，它们覆盖 W，且 $W_1 \cap W_2 = f^{-1}(V)$。集合层面的范坎彭等价（顶面是推出 $\Leftrightarrow$ 前面/右面是拉回）可提升至 $\mathrm{Def}_c(M)$：唯一的集合论中介映射是连续的（在开覆盖 $\{W_1, W_2\}$ 上应用粘贴引理），也是可定义的（在可定义开覆盖上分片可定义）。

证明处理二元情形$n = 2$；由归纳法可得n 个子对象的一般情形，每一步取$U_1 \cup \cdots \cup U_{n-1}$和$U_n$作为两个部分（中间的并是$X$的可定义开子集，因此是一个 $\mathscr{M}$-子对象）。

□

备注 A.5。$\mathrm{Def}_c(M)$ 不具备完全的 $\mathscr{M}$-粘合性：沿着一个 $\mathscr{M}$-单态射 $m : V \hookrightarrow U_1$ 作出的推出，若另一条边是任意连续的可定义映射 $f : V \to U_2$，未必满足 VK。一个反例：$V = (-1, 1)$，$U_1 = \mathbb{R}$，$U_2 = \mathbb{R}^2$，$m : V \hookrightarrow U_1$ 为包含映射，$f : V \to U_2$ 定义为 $f(t) = (t^2 - 1, t(t^2 - 1))$（节点三次曲线参数化；f 在 V 上是单射，但不是开浸入）。当 $t \to \pm 1$时，$f(t) \to (0, 0)$；在推出 $P = U_1 \sqcup_V U_2$, 中，任何从 P 出发的连续映射都必须将 $\mathrm{can}_1(\pm 1)$ 和 $\mathrm{can}_2(0, 0)$ 映到同一点，因此豪斯多夫商会将它们识别。于是拉回 $U_1 \times_P U_2$是 $[-1, 1]$，而不是 $(-1, 1) = V$：该推出方形不是拉回。性质 (A') 通过要求两条边都是 $\mathscr{M}$-单态射来避免这一问题。

B Sys 的粘合性$\mathrm{ho}(I, O)$

粘合性证明（定理 *2.8*）。设 $\mathscr{C}$ 为具有有限积的黏着的范畴，并设 U：$\mathrm{Sys}_{\mathrm{ho}}(I, O) \to \mathscr{C}$ 为遗忘函子 $(S\ \alpha) \mapsto S$。

拉回。U 构造拉回：给定 $(S1,\ \alpha_1) \xrightarrow{f} (S3,\ \alpha_3) \xleftarrow{g} (S2,\ \alpha_2)$，拉回状态空间 $P = S_1 \times_{S_3} S_2$ 在 $\mathscr{C}$ 中获得唯一的动力学 α_P：$P \times I \to P \times O$，使投影成为态射，因为 $- \times I$和 $- \times O$保持拉回。

沿单态射的推出。设$m\colon (C\ \gamma)\hookrightarrow(A\ \alpha_A)$是一个单态射（即$m\colon C\hookrightarrow A$在 $\mathscr{C}$中是单态射），且$f\colon (C\ \gamma)\to(B\ \alpha_B)$ 是一个态射。根据粘合性，推出 $P=A\sqcup_C B$在 $\mathscr{C}$ 中存在。动力学 $\alpha_P\colon P\times I\to P\times O$利用典范同一来定义

$$P\times I\ \cong\ (A\times I)\sqcup_{C\times I}(B\times I),$$

这由应用于如下立方体的范坎彭条件推出：其底面是推出 $A\sqcup_C B$，竖直映射是积投影 $-\times I\to -$（四个竖直面都是拉回：任意投影 $\pi_Y\colon Y\times Z\to Y$ 沿任意 $g\colon X\to Y$ 的拉回都是 $X\times Z$）。在 $A\times I$ 分量上，α_P是 $(\mathrm{can}_A\times \mathrm{id}_O)\circ\alpha_A$；在 $B\times I$ 分量上，α_P 是 $(\mathrm{can}_B\times \mathrm{id}_O)\circ\alpha_B$。$C\times I$ 上的相容性由 m 和 f 的态射条件得出；从推出 $(A\times I)\sqcup_{C\times I}(B\times I)$ 出发的映射由其限制唯一确定，因此动力学也是唯一的。

范坎彭条件。给定 $\mathrm{Sys}_{\mathrm{ho}}(I,\ O)$ 中的一个交换立方体，其底面是沿单态射的推出，且后面和左面是拉回，应用 U：根据 $\mathscr{C}$ 的粘合性，底层立方体的顶面是推出，当且仅当前面和右面是拉回。由于U 既创建拉回（如上），也创建沿单态射的推出——分配律的识别会在任何此类推出上确定唯一的动力学——这一等价关系可提升至 $\mathrm{Sys}_{\mathrm{ho}}(I,\ O)$：“若”方向成立，因为 U 创建拉回；“仅若”方向成立，因为 $\mathrm{Sys}_{\mathrm{ho}}(I,\ O)$ 中的推出是 $\mathscr{C}$ 中配备了唯一相容动力学的推出。

□

这里不需要可指数化假设：整个论证都适用于非柯里化动力学$\alpha\colon S\times I\to S\times O$，且从未构造 $(-)^I$。不过，假设可指数化，则上述内容是一般性定理 2.9 的一个特例。

定理 *2.9* 的证明。令$U\colon \mathrm{Coalg}(H)\to\mathscr{C}$为遗忘函子。

拉回。U 保留拉回：给定余代数态射$f\colon (A\ \alpha)\to(C\ \gamma)$和$g\colon (B\ \beta)\to(C\ \gamma)$，底范畴 $\mathscr{C}$中的拉回 $P=A\times_C B$ 通过$H(P)\cong H(A)\times_{H(C)}H(B)$（利用$H$ 保留拉回）获得唯一的余代数结构 $\pi\colon P\to H(P)$。

单态射。由于U是忠实的，它反映单态射。反过来，若$f\colon (A\ \alpha)\to(B\ \beta)$是 $\mathrm{Coalg}(H)$ 中的单态射，则其在底范畴 $\mathscr{C}$中的核对 $(R\ \pi_1,\pi_2)$ 由拉回的构造提升为余代数核对，因此 $\pi_1=\pi_2$ 可得$U(f)$ 在 $\mathscr{C}$中是单态射。因此，$\mathrm{Coalg}(H)$ 中的单态射与 $\mathscr{C}$中的单态射相对应。

沿单态射的推出。设 $m\colon (C\ \gamma)\hookrightarrow(A\ \alpha_A)$ 是一个单态射，且 $f\colon (C\ \gamma)\to(B\ \alpha_B)$ 是一个态射。由 $\mathscr{C}$ 的粘合性，推出 $P=A\sqcup_C B$ 存在。余极限余锥 α_A,α_B 通过泛性质在 P 上诱导出余代数结构：余锥 $D(j)\xrightarrow{\alpha_j}H(D(j))\ H(\iota_j)\longrightarrow H(P)$ 由 m 和 f 所满足的余代数态射条件而相容，因此泛性质给出 $\alpha_P\colon P\to H(P)$。唯一性：从推出出发的映射由其在 A 和 B 上的限制所确定，而余代数态射条件确定了这些限制。

范坎彭条件。给定 $\mathrm{Coalg}(H)$ 中的一个交换立方体，其底面是沿单态射的推出，且后面和左面是拉回，应用 U：由 $\mathscr{C}$ 的粘合性，底层立方体的顶面是推出，当且仅当前面和右面是拉回。这一等价性提升到 $\mathrm{Coalg}(H)$：“如果”方向使用拉回的构造，“仅当”方向则利用 $\mathrm{Coalg}(H)$ 中的推出映射到 $\mathscr{C}$ 中的推出，并配备唯一相容的余代数结构。

□

对于 $\mathscr{C}=\mathbf{Set}$，Padberg [19] 已在 $\mathscr{M}$-黏合范畴框架中证明了这一结果；对任意粘合基底成立的一般性结论似乎是新的。我们所用的拉回保持性质能否弱化为弱拉回（这将涵盖更多余代数相关函子 [13]），仍有待探究。

C 定理 5.1 的证明

证明。 证明利用了 Def(M) 的粘合性（§2.2），以及构造中的所有系统共享同一接口这一事实（I', O'）。

步骤 1：选择代表元和重叠见证。从每个等价类中选择代表元 $(\mathscr{S}'_\alpha, \psi_\alpha)$。交叠上的相容性意味着：对于每一对 (α,β)，限制 $(\mathscr{S}'_\alpha, \psi_\alpha \circ \mathrm{n}\alpha)$ 和 $(\mathscr{S}'_\beta, \Psi_\beta \circ \mathrm{n}\beta)$ 在 $\mathscr{F}\mathrm{j}(\mathscr{S}_\alpha \cap \mathscr{S}_\beta)$ 中芽等价，其中 $\mathrm{n}\alpha$: $\mathscr{S}_\alpha \cap \mathscr{S}_\beta \hookrightarrow \mathscr{S}_\alpha$。因此，存在一个系统 $\mathscr{S}''_{\alpha\beta}$，其接口为 (I′, O′)，以及单态射 $\mathrm{i}\alpha$: $\mathscr{S}''_{\alpha\beta} \hookrightarrow \mathscr{S}'_\alpha$、$\mathrm{i}\beta$: $\mathscr{S}''_{\alpha\beta} \hookrightarrow \mathscr{S}'_\beta$，它们具有恒等接口分量；还有一个态射 $\varphi_{\alpha\beta}$: $\mathscr{S}_\alpha \cap \mathscr{S}_\beta \to \mathscr{S}''_{\alpha\beta}$，满足 $\mathrm{i}\alpha \circ \varphi_{\alpha\beta} = \psi_\alpha \circ \mathrm{n}\alpha$ 和 $\mathrm{i}\beta \circ \varphi_{\alpha\beta} = \psi_\beta \circ \mathrm{n}\beta$。

步骤 *2*：通过余极限构造全局目标。将系统 $\mathscr{S}'$ 构造为图表的余极限

$$\cdots \mathscr{S}'_\alpha \xleftarrow{i_\alpha} \mathscr{S}''_{\alpha\beta} \xrightarrow{i_\beta} \mathscr{S}'_\beta \cdots$$

在 Sys 中。具体而言，$\mathrm{S}'_b = \mathrm{colim}(\mathrm{S}'_{\alpha,b} \leftarrow \mathrm{S}''_{\alpha\beta,b} \rightarrow \mathrm{S}'_{\beta,b})$，$\mathrm{S}'_a$也同理；由于所有 $\mathscr{S}'_\alpha$ 和 $\mathscr{S}''_{\alpha\beta}$都是齐次的，且所有态射都属于 $\mathrm{Sys}_{\mathrm{ho}}(\mathrm{I}', \mathrm{O}')$，因此两个余极限图表相同，所以 $\mathrm{S}'_b = \mathrm{S}'_a$且 $\mathscr{S}'$ 是齐次的。 $\mathscr{S}'$ 的接口为 (I′, O′)，因为所有部分共享此接口，且所有识别在接口上都是恒等映射。余极限存在于 Def(M) 中：它是 $\bigsqcup_\alpha \mathrm{S}'_{\alpha,b}$ 的商，由可定义等价关系生成；当 $\mathrm{s} \sim \mathrm{s}'$，且对某个 $\mathrm{t} \in \mathrm{S}''_{\alpha\beta,b}$，有 $\mathrm{s} = (\mathrm{i}\alpha)_b(\mathrm{t})$ 和 $\mathrm{s}' = (\mathrm{i}\beta)_b(\mathrm{t})$ 时，将二者等同。动力学相容性通过该等价关系的生成元传递：若 s 与 s′ 通过 $\mathrm{t} \in \mathrm{S}''_{\alpha\beta,b}$ 等同，则它们在动力学作用下的像也通过同一个 t 等同（因为$\mathscr{S}''_{\alpha\beta}$是一个子系统），因此根据链长归纳可知，该关系与动力学相容。

由于覆盖是有限的，这个余极限可以通过沿单态射迭代进行二元推出来构造。根据粘合性，每个沿单态射进行的二元推出都会保持单态射，因此归纳可得，典范映射 $\mathrm{can}\alpha$: $\mathscr{S}'_\alpha \hookrightarrow \mathscr{S}'$ 是具有恒等接口分量的单态射。

商上的动力学 α': $\mathrm{S}'_b \times \mathrm{I}' \to \mathrm{S}'_a \times \mathrm{O}'$ 定义良好：如果 $\mathrm{s} = (\mathrm{i}\alpha)_b(\mathrm{t})$ 与 $\mathrm{s}' = (\mathrm{i}\beta)_b(\mathrm{t})$ 对于 $\mathrm{t} \in \mathrm{S}''_{\alpha\beta,b}$而言被视为等同，那么 $\alpha'_\alpha(\mathrm{s}, \mathrm{i}')$ 和 $\alpha'_\beta(\mathrm{s}', \mathrm{i}')$ 在 S'_a中根据同一等价关系被视为等同，因为 $\mathscr{S}''_{\alpha\beta}$是 $\mathscr{S}'_\alpha$和 $\mathscr{S}'_\beta$的子系统。

步骤 *3*：粘合这些态射。这些复合 $\mathrm{can}_\alpha \circ \psi_\alpha$: $\mathscr{S}_\alpha \to \mathscr{S}'$ 构成一个相容族：对于$x \in \mathscr{S}_\alpha \cap \mathscr{S}_\beta$,

$$(\mathrm{can}_\alpha \circ \psi_\alpha)(n_\alpha(x)) = \mathrm{can}_\alpha(i_\alpha(\varphi_{\alpha\beta}(x))) = \mathrm{can}_\beta(i_\beta(\varphi_{\alpha\beta}(x))) = (\mathrm{can}_\beta \circ \psi_\beta)(n_\beta(x)),$$

中间的等式由余极限的定义性质得出。由于该覆盖在 $S_b \times I$和 $S_a \times O$上联合满射，相容族定义了一个唯一态射 ψ: $\mathscr{S} \to \mathscr{S}'$: 相容性确保了交叠处的良定义性，而动力学交换图在每个局部片上都成立，因此在整体上也成立。

步骤 *4*：验证接口与限制。 在每个 $\mathscr{S}_\alpha$, $\psi_I|_{I_\alpha} = (\mathrm{can}\alpha)_I \circ (j_I \circ m_{I,\alpha}) = j_I \circ m_{I,\alpha}$，因为 $\mathrm{can}\alpha$具有恒等接口。联合满射性给出 $\psi_I = j_I$；类似地， $\psi_O = j_O$。因此 $(\mathscr{S}', \psi) \in \mathscr{F}_j(\mathscr{S})$。

对 $\mathscr{S}_\alpha$ 的限制满足 $\psi \circ m_\alpha = \mathrm{can}_\alpha \circ \psi_\alpha$。由于 $\mathrm{can}\alpha$: $\mathscr{S}'_\alpha \hookrightarrow \mathscr{S}'$ 是具有恒等接口的单态射，取 $\mathscr{S}'' = \mathscr{S}'_\alpha$, $i_1 = \mathrm{can}\alpha$, $i_2 = \mathrm{id}$, $\varphi = \psi_\alpha$ 可作为 $\mathscr{F}_j(\mathscr{S}_\alpha)$ 中 $(\mathscr{S}', \mathrm{can}_\alpha \circ \psi_\alpha) \sim (\mathscr{S}'_\alpha, \psi_\alpha)$ 的见证。 □

D 定理 5.5 的证明

证明。（⇒）：稳健不连通的 ⇒不是层。假设$j_I^{-1}(t_0)$稳健不连通，且$j_I^{-1}(N) \subseteq V \cup W$。取一个闭可定义的 $\bar{}N_0 \subset N$，使$t_0 \in \mathrm{int}(\bar{}N_0)$，并令$C := I \setminus j_I^{-1}(\bar{}N_0)$。由于

$\mathrm{im}(\mathrm{j}_O)$ 是由 ≥ 2 个点构成的可定义连通集，O 的某个可定义连通分支 C_0 上 $\mathrm{j}_O|_{C_0}$ 非常值。取 $\mathrm{o}_0,\mathrm{o}_1 \in \mathrm{C}_0$，使 $\mathrm{j}_O(\mathrm{o}_0) \neq \mathrm{j}_O(\mathrm{o}_1)$，并用一条可定义弧 $\tilde{\gamma}\colon [0,1] \to \mathrm{C}_0$ 将它们连接起来，其中$\tilde{\gamma}(0) = \mathrm{o}0, \tilde{\gamma}(1) = \mathrm{o}_1,$；令 $\beta\colon \bar{\mathrm{N}}_0 \to [0,1]$ 为满足 $\beta|_{\partial \bar{\mathrm{N}}_0} = 0, \beta(\mathrm{t}_0) = 1$ 的可定义鼓包函数（存在性：紧致可定义集的正规性，参见 [6, 第 6] 章）。定义 $\mathrm{f}\colon \mathrm{I} \to \mathrm{O}$：在 $\mathrm{V} \cap \mathrm{j}_I^{-1}(\bar{\mathrm{N}}_0)$ 和 $\mathrm{I} \setminus \mathrm{j}_I^{-1}(\bar{\mathrm{N}}_0)$ 上，令 $\mathrm{f}(\mathrm{x}) = \mathrm{o}_0$；在 $\mathrm{W} \cap \mathrm{j}_I^{-1}(\bar{\mathrm{N}}_0)$ 上，令 $\mathrm{f}(\mathrm{x}) = \tilde{\gamma}(\beta(\mathrm{j}_I(\mathrm{x})))$。于是 f 连续（V 和 W 部分在 $\mathrm{j}_I^{-1}(\bar{\mathrm{N}}_0)$ 中既开又闭，且 $\mathrm{f} \to \mathrm{o}_0$ 在边界处）。在 U_1, 上，f 为常值 o_0,，因此 $\mathrm{j}_O \circ \mathrm{f}|_{U_1}$ 可经 j_I 因子分解。在 U_2, 上，$\mathrm{W} \cap \mathrm{j}_I^{-1}(\bar{\mathrm{N}}_0)$ 部分仅取决于 $\mathrm{j}_I(\mathrm{x})$，C 部分为常值，并且 $\mathrm{W} \cap \mathrm{j}_I^{-1}(\bar{\mathrm{N}}_0)$ 中没有任何点与 C 中的点具有相同的 j_I 值，因此因子分解成立。覆盖 $\mathrm{U}_1 := \mathrm{V} \cup \mathrm{C}, \mathrm{U}_2 := \mathrm{W} \cup \mathrm{C}$ 给出两个截面 $\mathrm{j}_O \circ \mathrm{f}|_{U_i}$，它们在 $\mathrm{U}_1 \cap \mathrm{U}_2 = \mathrm{C}$ 上相同，但对于 $\mathrm{a} \in \mathrm{V} \cap \mathrm{j}_I^{-1}(\mathrm{t}_0)$、$\mathrm{b} \in \mathrm{W} \cap \mathrm{j}_I^{-1}(\mathrm{t}_0)$，满足 $\mathrm{j}_O(\mathrm{f}(\mathrm{a})) = \mathrm{j}_O(\mathrm{o}_0) \neq \mathrm{j}_O(\mathrm{o}_1) = \mathrm{j}_O(\mathrm{f}(\mathrm{b}))$。

($\Leftarrow$)：无稳健不连通 $\Rightarrow$ 层。设 $\mathrm{h}\colon \mathrm{I} \to \mathrm{O}'$ 在某个覆盖$\{\mathrm{I}_\alpha\}$上局部保持 $\ker(\mathrm{j}_I)$不变，但并非全局如此：存在 x、y，使得 $\mathrm{j}_I(\mathrm{x}) = \mathrm{j}_I(\mathrm{y}) = \mathrm{t}_0$，且 $\mathrm{h}(\mathrm{x}) \neq \mathrm{h}(\mathrm{y})$。由于对每个 α，h 在 $\mathrm{j}_I^{-1}(\mathrm{t}_0) \cap \mathrm{I}_\alpha$ 上为常值，函数 $\mathrm{h}|_{j_I^{-1}(\mathrm{t}_0)}$ 是局部常值函数，其水平集既开又闭。令 $\mathrm{A} := \{\mathrm{z} \in \mathrm{j}_I^{-1}(\mathrm{t}_0) : \mathrm{h}(\mathrm{z}) = \mathrm{h}(\mathrm{x})\}$，$\mathrm{B} := \mathrm{j}_I^{-1}(\mathrm{t}_0) \setminus \mathrm{A}$（两者均非空、互不相交，且在 I 中闭）。由 I 的正规性（I 是紧致豪斯多夫空间），存在互不相交的开集 $\mathrm{V} \supseteq \mathrm{A}$、$\mathrm{W} \supseteq \mathrm{B}$。由于 $\mathrm{S} := \mathrm{I} \setminus (\mathrm{V} \cup \mathrm{W})$ 是闭集，且 $\mathrm{S} \cap \mathrm{j}_I^{-1}(\mathrm{t}_0) = \varnothing$，由紧致性可得 $\mathrm{t}_0 \notin \mathrm{j}_I(\mathrm{S})$，因此存在 t_0 的某个开邻域 $\mathrm{N} \ni$，使得 $\mathrm{j}_I^{-1}(\mathrm{N}) \subseteq \mathrm{V} \cup \mathrm{W}$ ——这就构成了稳健不连通。既然每个局部保持 $\ker(\mathrm{j}_I)$不变但并非全局保持不变的函数都会导致稳健不连通，逆否命题表明：不存在稳健不连通，就意味着每个局部不变函数都是全局不变的，即 $\mathscr{F}_j^{\mathrm{ri}}$,1 是一个层。 □

E 反例图

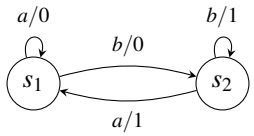


图 1：反例 4.4 中的双状态系统（对 $\mathscr{F}_j^{\mathrm{ri}}$ 的分离失败）。边标签：输入/输出。输入分割覆盖 $\mathscr{S}|_{\{a\}}$, $\mathscr{S}|_{\{b\}}$ 将每个局部片限制为单一输入；另一种解释在纯输入序列上保持一致，但在混合序列（a，b）上不一致。

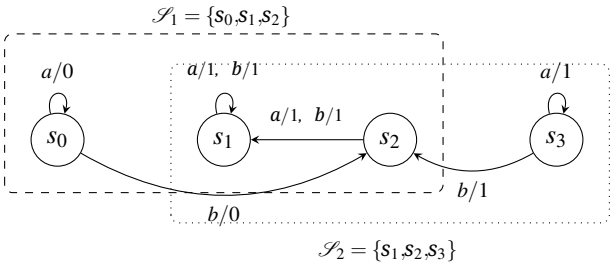


图2：反例 5.2 的四状态系统（ $\mathscr{F}_j^{\text{beh}}$ 的粘合失败）。边标签：输入/输出；判定器将输入合并（j_I： $\{a,\ b\} \to \{\bullet\}$）。虚线框和点线框表示覆盖补丁；交叠为 $\{s_1, s_2\}$。阻碍：$s_0 \xrightarrow{b} s_2$ 和 $s_3 \xrightarrow{b} s_2$ 对后继状态s_2提出了不相容的行为要求。

F 符号索引

范畴与空间

$\mathscr{C}$ 基范畴（通常是黏着的且具有有限积） Def(M) o-极小结构 M 中可定义集合的范畴 Def_c(M) Def(M) 的宽子范畴：连续可定义映射 Coalg(H) 自函子 H 的余代数范畴 Sys 异构Mealy机（接口可变）的范畴 Sys_{ho}(I, O) 非满子范畴：齐次系统， $\text{f}_b = \text{f}_a\text{Sys}_{\text{oi}}$ Sys 的宽子范畴：可定义开浸入 Set 集合范畴 Systems (S_b, S_a, I, O α) 异构Mealy机（系统）α： $\text{S}_b \times \text{I} \to \text{S}_a \times \text{O}$ 动力学映射ι： $\mathscr{S}' \hookrightarrow \mathscr{S}$ 系统的可定义开浸入 判定器与接口 $\text{j} = (\text{j}_I, \text{j}_O)$ 判定器：到可解释接口 (I′, O′) 的态射 I'_U 子系统的受限接口 $\text{im}(\text{j}_I \circ \text{m}_I)$ $\mathscr{S}_U$预层与等价关系$\mathscr{F}_j$ 余芽商解释预层 $\mathscr{F}$behj 行为预层（按 $\approx$取商；分离）$\mathscr{F}$rij 受限接口预层$\mathscr{F}$rij,1 $\mathscr{F}_j^{\text{ri}}$的无状态子预层（截面满足 $|\text{S}'| = 1$）$\sim$ 截面上的余芽等价$\approx$ 截面上的行为等价