arXiv:2603.07268v1 [cs.LO] 7 Mar 2026

面向草图的数据库

Dominique Duval* Rachid Echahed†

摘要

本文介绍了面向草图的数据库，这是一种范畴框架，将数据库范式编码为有限极限图式，并将单个数据库和模式表示为集合值模型。本文通过箭图、RDF 三元组存储和属性图等面向图的范式，展示了这一形式体系。本文还说明，标签、属性、类型化和路径等常见图特征如何通过草图构造得到统一刻画。由于路径在查询中发挥重要作用，我们提出了通过局部化器形式化的推理规则，以便按需计算有用路径；这类局部化器也适用于数据库类型符合性等任务。最后，本文介绍了停顿图式，旨在促进模块化组合和可扩展的模型增长：停顿图式是有限极限图式，其中关系由单个极限而非两个嵌套极限指定；本文还证明，停顿图式的模型的有限并是逐点余极限。

关键词： 图式、范畴论、数据库

1 引言

数据管理系统的发展经历了显著转变：从僵化、以表为中心的关系范式[1] [8, 9] 转向图数据库等表达能力更强、更灵活的范式。这些系统包括资源描述框架（RDF）图或三元组存储 [26, 25] ，以及属性图 [1, 11], ，旨在捕捉现代数据中固有的复杂关系和语义结构。然而，尽管在实践中取得了成功，不同的图数据库范式仍缺乏统一的形式化基础，无法支持严谨推理、组合语义和有原则的推断。本文提出一个适用于多种数据库范式（包括面向图的范式）的范畴框架，该框架基于有限极限图式 [5, 4], ，为数据库设计、语义和推断提供一种统一且富有表达力的模型论方法。

图数据库通常由节点和边组成，其中边表示节点之间的二元关系。这些边可以拼接成路径，从而表示组合关系并支持导航式查询。用范畴论术语来说，这种结构类似于一个小范畴，其中对象对应节点，态射对应路径。然而，现实世界中的图数据库还包含属性-值对、标签和类型方案等额外特性，这些特性超出了基本的范畴结构。

自 20 世纪 90 年代初以来，文献中一直在探讨范畴论在数据库系统各个方面的应用；代表性研究包括 [15, 19, 23, 2, 21, 24, 20, 27]。这些

*LJK，格勒诺布尔-阿尔卑斯大学，法国，dominique.duval@univ-grenoble-alpes.fr

†LIG、法国国家科学研究中心和格勒诺布尔-阿尔卑斯大学，法国，rachid.echahed@imag.fr[1]我们使用 数据库范式 而不是 数据库模型 ，以避免与论文中讨论的其他模型相关概念混淆。

研究主要关注数据库语义——模式、数据库及其间的关系（类型化、转换及相关概念）。关系范式受到了最多关注，而实体—关系（ER）、面向对象数据库以及面向图的模型（例如资源描述框架（RDF））等其他范式受到的关注则较少。

基于这些范畴基础，已有多位作者采用草图的概念[5, 4], ——一种植根于范畴论的图示形式体系——来刻画数据库范式的不同方面。典型示例包括 [7, 12, 13, 18], 将草图用作一种灵活语言，以对不同范式中的结构属性和语义属性进行建模。

多范式数据库系统的出现 [17], ——这类系统在单一后端中集成多个数据库，并将其作为可能属于不同范式的模型——使范畴工具和草图尤为重要。近期研究将范畴论构造应用于多范式场景，引入模式范畴和数据库范畴等概念，以关联异构数据库 [22, 16, 10]。这些方法拓展了早期研究：范畴不仅用于描述模式和数据库，也用于阐明不同范式之间的关系。

与这些先前的研究不同，我们采取了更具基础性的立场，将视角提升一个层次。我们不直接刻画模式或数据库，而是从形式化范式本身入手。随后，模式和数据库被视为这些范式的模型。为此，我们采用草图这一句法框架，具体而言，采用有限极限草图 [5, 4], 作为元语言，来刻画关系范式、RDF 三元组存储范式和属性图范式等核心数据库范式，从而以范畴论为其结构和行为奠定基础 [14, 6]。

有限极限草图借助一个由节点和边构成、并辅以指定的环、三角形和有限锥加以充实的箭图，来呈现配备了选定有限极限（积、拉回、等化子等）的范畴。一个草图生成一个理论（具有有限极限的范畴）以及一个集合值模型范畴（保持这些极限的函子）。例如，箭图的草图刻画了节点与边的基本结构，而对该草图的扩展则可以表示标记箭图、三元组存储和属性图。将数据库范式解释为有限极限图式、将单个数据库解释为该图式的模型，会带来若干实际优势：模型范畴中表现良好的极限与余极限、由示意图态射诱导的典范伴随，以及可直接运用米田引理来关联句法与语义。这些范畴工具能够自然地刻画常见的数据库构造（类型化、属性、路径生成），同时使其语义明确且可复合。例如，RDF 三元组存储可以建模为边由其源、目标和标签所确定的标记箭图。属性图包含属性-值对和类型化信息，可通过扩展前述草图、增加表示属性和模式约束的额外的点与锥来建模。关系数据库中的表仅仅由单元格构成，单元格即（行，列）对，而表的内容可以看作从单元格到值的偏函数。这些构造保持了该范式以图式为导向的性质，从而确保所有数据库特征都能以范畴方式加以刻画。

除了结构建模之外，我们的框架还通过局部化器这一概念支持推理。局部化器是一种图式态射，是范畴论中局部化概念在图式中的对应物。其思想是区分“理想化”数据库（包含所有路径、每个实体的所有可能类型等）和更“具体”的数据库（采用“惰性”方式构造路径等）。每个具体数据库都是理想化数据库的表示。推理过程用于修改表示，而不改变其所表示的理想化数据库，也因此不改变其语义。面向图式的推理系统与局部化器相关联，其推理规则是一些箭头，这些箭头经局部化器作用后会变为可逆箭头。应用推理规则对应于具体数据库范畴中的推出。这种方法是逻辑推理的一种替代方案，并支持针对路径、类型层次结构和语义约束进行推理。

为了解决可扩展性和组合性问题，我们引入停顿图式，这是一类新型有限极限图式，旨在控制模型大小并支持模型的逐点并。停顿图式简化了关系的表示。这项创新对于管理大型图尤为重要。我们证明，停顿图式的模型的有限并是逐点余极限，从而确保组合语义和易处理推理。

本文始终强调草图所提供的范畴基础和语义清晰性，并主张采用草图作为基于图的数据系统及其他系统的统一框架。第 2 节将面向草图的数据库设定形式化，并展示其在各种图数据库范式中的应用。第 3 节介绍面向草图的推理系统，定义局部化器，并将其应用于路径构造。第 4 节介绍停顿图式，确立其理论性质，并探讨其在模型复合中的作用。第 5 节给出结语。我们假设读者熟悉范畴论（例如，参见[14, 6]了解缺失的概念）。附录 A 新增了关系数据库范式的基于草图的表示示例。

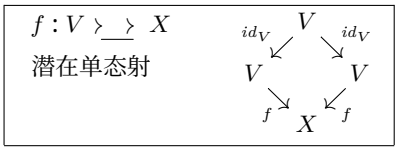
2 以图式为导向的数据库

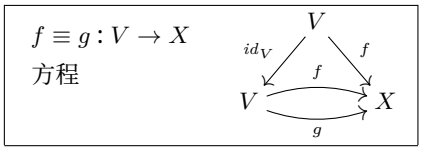
本节旨在回顾有限极限图式及其模型的定义，并通过示例说明，有限极限图式可以视为一种数据库范式，而该图式的模型则可视为相应的数据库。第 2.1 节回顾一些范畴工具，第 2.2 节说明这些工具在数据库中的应用，第 2.3 节则将其应用于若干图数据库范式。

2.1 有限极限草图

我们依据 [5] 介绍有限极限草图，并对术语以及方程的引入方式做了一些小幅修改。有限极限草图是具有有限极限的范畴的表示。它们由称为箭图的简单图类型定义。箭图$\mathbb{Q}$ 由一个点（或宾语）集合和一个箭头集合组成，并配有从箭头到点的两个函数：定义域和陪域。一个定义域为 X 、陪域为 Y 的箭头 f 记作 $f\colon\ X \to Y$。箭图态射由一个作用于点的函数和一个作用于箭头的函数组成，并且这两个函数分别与定义域和陪域相容。在箭图 $\mathbb{Q}$中，环是一个箭头 $f\colon X \to X$ ，三角形由三个箭头组成：$f\colon\ X \to Y$、$g\colon\ Y \to Z$和 $h\colon\ X \to Z$。一个形状为 J的图表 D（其中 J是一个箭图）是从 J到 $\mathbb{Q}$ 的箭图态射。一个锥 C 由一个称为C的基的图表 $D\colon\ J \to \mathbb{Q}$ 、一个称为 C 的顶点的点 V（属于 $\mathbb{Q}$），以及对于 J中的所有 X，一组来自 $\mathbb{Q}$的投影 $p_X\colon\ V \to D(X)$组成，这些投影称为 C的投影。当 J 为有限集时，它是一个有限锥。在此阶段，不假定交换性。

一个有限极限图式 $\mathbb{S}$ 由一个箭图 $\mathbb{Q}$构成，称为 $\mathbb{S}$的底层箭图 ，并带有 $\mathbb{Q}$的一组环，称为 $\mathbb{S}$的潜在恒等式 ， $\mathbb{Q}$的一组三角形，称为 $\mathbb{S}$的潜在复合 ，以及 $\mathbb{Q}$的一组有限锥，称为 $\mathbb{S}$的潜在有限极限 。我们使用与范畴中相同的记号（id_X 和 $g \circ f$）来表示潜在恒等式和潜在复合。这里的“潜在”一词意味着：这些在草图中并无特定属性的特征，在每个模型（定义如下）中会成为“实际”特征。例如，一个潜在恒等式在草图中只是一个环，而在每个模型中会成为恒等函数。请注意，“潜在极限”也可能有其他名称，如“形式极限”或“指定锥”。例如，下面的潜在极限及其简写记号分别表示：箭头 f 是一个潜在单态射，而 $f \equiv g$ 是一个潜在等式，也称为方程。





在集合中，关系是笛卡尔积的子集。这个概念可推广如下。在一个范畴中，图上的关系 D 由两个连续的极限构成：首先是 D 的极限，然后是一个陪域为该极限顶点的单态射。请注意，点上的关系 X 只是一个陪域为 X 的单态射。在图式中也有类似的定义，其中包含两个连续的潜在极限。

有限极限草图的态射是保持所有潜在特征的箭图态射。

有限极限图式 $\mathbb{S}$ 的理论 Th($\mathbb{S}$) 是由$\mathbb{S}$ 自由生成的具有有限极限的范畴，其中 $\mathbb{S}$ 的所有潜在特征都成为实际特征。有限极限图式之间的每个态射 $\mathsf{s} : \mathbb{S} \to \mathbb{S}'$ 都会生成一个保有限极限函子 $\mathsf{Th(s)} : \mathsf{Th}(\mathbb{S}) \to \mathsf{Th}(\mathbb{S}')$。当函子 Th(s) 是范畴等价时，称态射 s 为等价；由此生成的等价关系称为图式等价。例如，在给定图表上添加一个极限就是图式等价。有限极限图式 $\mathbb{S}$ 的一个模型 M（本文中的模型均为集合值模型，除非另有明确说明）为 $\mathbb{S}$ 的每个点 X 指定一个集合 $M(X)$，并为 $\mathbb{S}$ 的每个箭头 $f\colon X \to Y$ 指定一个函数 $M(f) : M(X) \to M(Y)$，使得 $\mathbb{S}$ 的每个潜在特征都成为集合范畴中的实际特征。 $M(X)$中的元素称为 M 类型为 X 的元素。 $\mathbb{S}$ 的模型之间的态射 $m : M \to M'$ 是一个函数族$m_X\colon M(X) \to M'(X)$，对 $\mathbb{S}$ 中的每个点 X 都有定义，并且对于 $\mathbb{S}$ 中的每个箭头$f : X \to Y$，满足 $m_Y \circ M(f) = M'(f) \circ m_X$。这定义了模型范畴 Mod($\mathbb{S}$)，它等价于从 Th($\mathbb{S}$) 到集合范畴的保有限极限函子所构成的范畴，其中态射为自然变换。因此，等价的有限极限图式的模型范畴彼此范畴等价。

显然，每个有限极限图式的态射 $\mathsf{s} : \mathbb{R} \to \mathbb{S}$ 都会确定一个函子 $U_{\mathsf{s}} : \text{模型}(\mathbb{S}) \to \text{模型}(\mathbb{R})$，使得对于 $\mathbb{S}$的每个模型 M， $\mathbb{R}$的模型 $U_{\mathsf{s}}(M)$ 定义为 $U_{\mathsf{s}}(M)(\text{–}) = M(\mathsf{s}(\text{–}))$。有限极限图式的一个重要属性是，函子 U_{s} 有一个左伴随 $F_{\mathsf{s}} : \text{模型}(\mathbb{R}) \to \text{模型}(\mathbb{S})$。我们分别将 U_{s} 和 F_{s} 称为与 s 相关联的**底函子** 和**自由函子** 。有限极限图式 $\mathbb{S}$ 的另一个重要属性是，范畴 模型($\mathbb{S}$) 中存在所有极限和余极限。此外，模型($\mathbb{S}$) 中的所有极限都是逐点的，因此可以通过集合的极限轻松计算。当 $\mathbb{S}$ 只是一个箭图时，模型($\mathbb{S}$) 中的余极限是逐点的，但一般而言并非如此。模型余极限的重要性及其计算方法将在第 3 节和第 4 节讨论。请注意，有限极限图式及其模型范畴有多个名称，反映了不同的观点：例如，有限极限图式对应于本质代数规范 ，而它们的模型范畴就是局部有限可呈范畴。关于有限极限图式的模型与某种广义霍恩理论的模型之间的比较，可见 [3]。此外，图式的定义还有许多变体。例如，图式既可以基于箭图定义，也可以基于范畴定义。大多数图式定义都使用图表来指定等式，而我们为此使用部分复合和等化子，它们都是有限极限。这样一来，只需删去“潜在”一词，就能简单地定义图式的理论及其模型。不过，也可以采用其他任何图式定义。

将图式态射的概念推广为基于范畴分式的“等价意义下的态射”概念，往往是有必要的，我们称之为“多态映射”[2]。从草图 $\mathbb{R}$ 到草图 $\mathbb{S}$ 的多态映射 s 由一个草图 $\mathbb{S}'$ 以及两个态射 $\mathsf{s}_n : \mathbb{R} \to \mathbb{S}'$ 和 $\mathsf{s}_d : \mathbb{S} \to \mathbb{S}'$构成，

[2]在生物学中，多态映射是某些微生物改变其形状或大小的能力。

分别称为 s 的分子 和分母 ，且 s_d 是一个图式等价。多态态射可以像范畴分式一样，借助图式的推出进行复合。每个态射 $\mathsf{s}:\mathbb{R}\to\mathbb{S}$ 都可视为一个多态映射，以 s 为分子、以 $\mathbb{S}$的恒等态射为分母。用于态射的记号 $\mathsf{s}:\mathbb{R}\to\mathbb{S}$ 被推广到多态态射。事实上，与图式态射一样，多态映射 $\mathsf{s}:\mathbb{R}\to\mathbb{S}$ 生成一个有限极限保持函子 $Th(\mathsf{s}):Th(\mathbb{R})\to Th(\mathbb{S})$，并在 $Mod(\mathbb{S})$ 与 $Mod(\mathbb{R})$ 之间确定一个伴随 $F_{\mathsf{s}}\dashv U_{\mathsf{s}}$。

在本文中，草图表示有限极限图式；若不会造成歧义，则可以省略“潜在的”一词。

2.2 面向图式的数据库

在本文中，我们将数据库范式定义为一个图式 $\mathbb{S}$ ，并将相对于此范式的数据库定义为 $\mathbb{S}$ 的一个模型。

巧合的是，图式和图数据库都是由箭图构成的。为了避免图式的“元层级”与数据库的“数据层级”之间产生混淆，我们对这一相同概念采用两套不同的词汇。在讨论数据库时，一个箭图由一组节点、一组边以及两个从边到节点的函数构成，分别称为源和目标。因此，箭图是以下图式的一个模型：

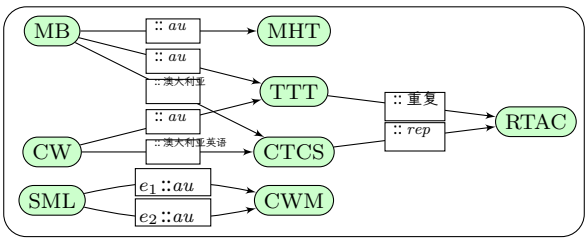
$$\mathbb{S}_q=\boxed{N\underset{t}{\overset{s}{\rightrightarrows}}E}$$

在图数据库的语境中，多个边可以具有相同的源和目标。此外，边通常还会关联第三种数据，通常是一个字符串，我们称之为标签。我们将标记箭图定义为每条边都关联一个标签的箭图。因此，标记箭图是以下图式的一个模型：

$$\mathbb{S}_{\ell q}=\boxed{N\underset{t}{\overset{s}{\rightrightarrows}}E\xrightarrow{u}L}$$

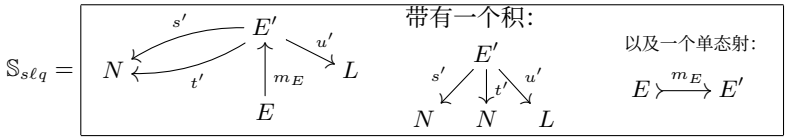
记号。 本文中，箭图中的节点 n 记为 (n)，源为 n_1 、目标为 n_2 的边 e 记为 $(n_1)\text{–}[e]\to(n_2)$。在标记箭图中，源为 n_1、目标为 n_2 、标签为 ℓ 的边 e 记为 $(n_1)\text{–}[e::\ell]\to(n_2)$。当从 n_1 到 n_2 仅有一条标签为 ℓ 的边时，该边称为强标记，此记号可简化为 $(n_1)\text{–}[::\ell]\to(n_2)$，或简记为 $(n_1)\text{–}\ell\to(n_2)$。

例 1。 我们构建一个玩具数据库，其预期含义如下：M. Barr（MB）是论文 [3] （MHT）的作者，M. Barr（MB）与 C. Wells（CW）是书籍 [5]（CTCS）和 [4] （TTT）的共同作者，这两本书转载于《范畴论的理论与应用重印本》（RTAC），而 S. Mac Lane（SML）是书籍 [14] （CWM）两个版本的作者。



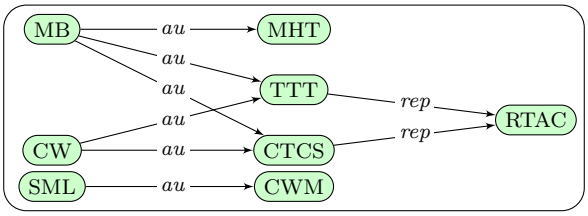
强标记箭图是一种标记箭图，其中每条边都带有强标记，也就是说，每条边都由其源、目标和标签组成的三元组唯一确定。这是一种

重载，因此多条边可以共享同一个标签。因此，强标记箭图是以下草图的一个模型:



记号。 如上所述，在强标记箭图中，边的记号 $(n_1)\ -[::\ell]\rightarrow(n_2)$ 被简化为 $(n_1)\text{–}\ell\text{→}(n_2)$。

例 2。 例 1 中的标记箭图在从 (SML) 到 (CWM) 的两条边合并后成为强标记箭图。$(n_1)\text{–}au\text{→}(n_2)$ 的预期含义是， n_1 是n_2的作者。



显然，每个标记箭图都是一个箭图，并且每个强标记箭图都是一个标记箭图。这可以看作分别与 $\mathbb{S}_q$在 $\mathbb{S}_{\ell q}$中的包含以及从 $\mathbb{S}_{\ell q}$到 $\mathbb{S}_{s\ell q}$的多态映射相关联的底函子的应用，该多态映射将 N, E, L, s, t, u映到$N, E, L, s' \circ m_E, t' \circ m_E, u' \circ m_E$。

2.3 一些以图式为导向的数据库范式

本节将考察几种图数据库范式，说明它们都可以视为以图式为导向。附录 A 说明，关系数据库范式也属于这种情况。

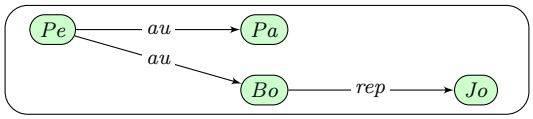
•RDF 图。

资源描述框架资源描述框架（或*RDF* 图，或*RDF* 三元组存储）对应的图是一个强标记箭图，对节点和边的性质有一些约束 [26]。它描述有关资源的陈述。资源对应一个节点，而陈述对应一条边，这条边称为*RDF* 三元组。RDF 三元组的标签称为谓词，其源和目标分别称为主语 和宾语。

•ER 图。

实体—关系图实体—关系图（或*ER*图）是一种强标记箭图，用于描述特定知识领域中彼此相关且值得关注的事物。它用于设计关系型或面向图的数据库。值得关注的事物称为实体，并按实体类型分类，这些类型对应于节点。两个节点之间的边表示相应实体类型实例之间的关系。边的标签是关系类型。

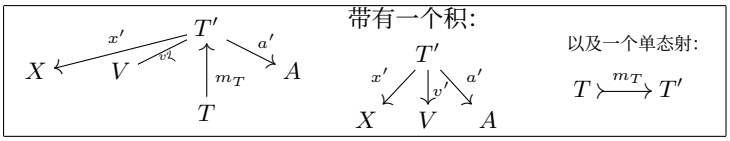
例3。 以下 ER图将作为例5中图数据库的模式。实体类型包括Pe （人物）、Pa （论文）、Bo （书籍）和Jo （期刊）。关系类型包括au （“是.....的作者”）和rep （“转载于”）。



• **属性-值对**。

从任意一种以图式为导向的数据库范式 $\mathbb{S}$ 以及 $\mathbb{S}$中的任意点 X 出发，我们验证，在 X 上添加属性-值对（也称为键值对或属性）后，该范式仍保有以图式为导向的性质。为此，将草图 $\mathbb{S}$ 扩展为包含一个表示属性的点 A、一个表示值的点 V、一个表示项-属性-值三元组的点 T，以及潜在极限，用于断言 T 是 X、 A 和 V上的关系。这意味着将 $\mathbb{S}$ 与以下草图$\mathbb{S}_{av,\ X}$ 在点 X处合并。

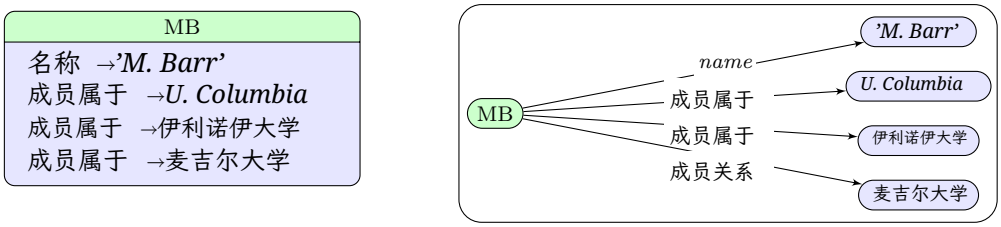
$\mathbb{S}_{av,X} =$



例如，将草图 $\mathbb{S}_{q}$, $\mathbb{S}_{av,\ N}$ 和 $\mathbb{S}_{av,\ E}$ 在 N, E, V 和 A上合并，我们就得到一个描述节点和边带有属性-值对的箭图的草图。这是构造属性图草图（定义1）的一步。

记号。 我们用 $a \to v$ 表示一个属性-值对。这种记号与以下事实相符：箭图的节点 n 上的每个属性-值对 $a \to v$ 都可以替换为一条强标记边（n）$-a\to$（v），前提是将值视为节点。事实上，在 $\mathbb{S}_{av,\ N}$ 中识别 N 和 V，便可得到一个与强标记箭图的草图 $\mathbb{S}_{s\ell q}$ 同构的草图。

例4。 本例说明了如何将节点上带有属性-值对的箭图（左侧）转换为带有强标记边的箭图（右侧）：



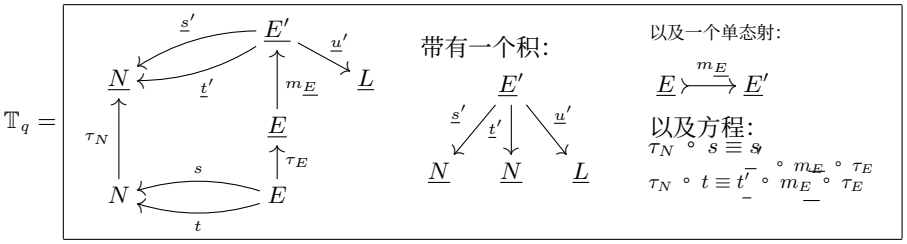
• **模式（唯一定型）**。

与关系数据库相比，图数据库的模式相当灵活：它们不是强制性的，可以不断演变，每个节点或边都可以有多种类型，而且类型之间存在层级关系。从任意一种草图导向数据库范式出发，我们会检查添加模式与类型化的概念是否能保留该范式的草图导向特性。这里我们先介绍一种简单的模式概念，其中每个元素恰好有一种类型；随后（在第 2.3 节）介绍一种更一般的概念，其中类型之间存在层级关系。

让我们从一个草图 $\mathbb{S}$ 开始，它规定了一些无类型数据库。最简单的情况下，无类型数据库 G 和模式 Sch 都是 $\mathbb{S}$ 的模型，而类型化是 $Mod(\mathbb{S})$ 中的一个态射 $\tau : G \to Sch$。然而，模式可能具有更丰富的结构，因此我们给出以下关于唯一定型的定义。给定一个多态映射 $\mathsf{s} : \mathbb{S} \to \mathbb{S}$，唯一定型的模式 是 $\mathbb{S}$ 的一个模型 Sch，而基于 Sch 的唯一类型化数据库 由 $\mathbb{S}$ 的一个模型 G 和一个态射 $\tau : G \to U_{\mathsf{s}}(Sch)$ 组成，该态射属于

$Mod(\mathbb{S})$。因此，令 $\mathsf{s}_n : \mathbb{S} \to \mathbb{S}'$ 表示 s 的分子，则唯一类型化数据库是草图 $\mathbb{T}$ 的模型；该草图由 $\mathbb{S}$的一个副本、 $\mathbb{S}'$的一个副本、对 $\mathbb{S}$ 中的每个点X，一个箭头 $\tau_X : X \to \mathsf{s}_n(X)$，以及对 $\mathbb{S}$中的每个箭头 $f : X \to Y$，一个方程 $\tau_Y \circ f \equiv \mathsf{s}_n(f) \circ \tau_X$ 组成。

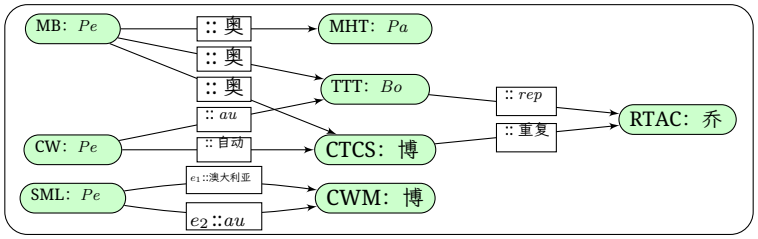
例如，从箭图出发，并认定模式是一个强标记箭图，我们将唯一类型化箭图 定义为由一个强标记箭图Sch、一个箭图 G，以及一个从 G 到Sch底层箭图的态射（称为类型化态射）构成。节点类型 和边类型是Sch的节点和边。然后，可以将 G 中任意边的标签 定义为其类型的标签，这使 G 成为一个标记箭图，而它一般并非强标记。因此，唯一类型化箭图是以下草图的模型：



注释 1. 此定义阐明了两个概念之间的相似之处和差异。本文分别将它们称为边类型 和标签，它们在各种图数据库中有着不同的名称。总而言之，边类型由源和目标（均为节点类型）以及其标签构成的三元组确定，而边的标签就是其类型的标签。请注意，类型化可应用于任何类型的数据库范式 $\mathbb{S}$ 以及 $\mathbb{S}$中的任意点 X ；而本文定义的标签仅适用于图数据库中的边。

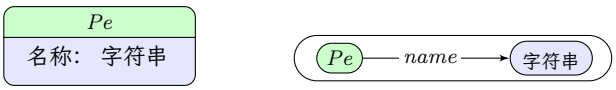
记号。 x 是 x 的类型这一事实记作 $x{:}\, x$。此外，每条带类型边（$n_1{:}\, \underline{n_1}$）$-[e{:}\, e]\to$（$n_2{:}\, n_2$）也可记作（$n_1{:}\, \underline{n_1}$）$-[\underline{e}{::}\, \ell]\to$（$\underline{n_2}{:}\, n_2$），其中 ℓ 是 e的标签，因为可从此记号中通过三元组（n_1）$-\ell\to$（n_2）还原e 。

例5. 例1中的无类型数据库现在已添加了例3中ER图的类型。

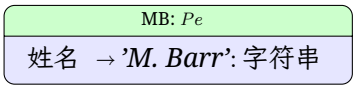


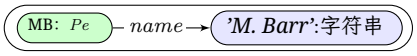
唯一定型很容易从箭图推广到节点和边上带有属性-值对的箭图，其中模式定义为一个强标记箭图，其节点和边上的属性 $a{:}\, a$ 具有类型化属性。因此，节点和边上带有属性-值对的唯一类型化箭图是某个草图的模型。

例6. 聚焦于单个节点，本例展示了节点上使用属性-值对记号的类型化概念（左侧）与强标记边的概念（右侧）相对应。模式如下：



以及节点上带有属性-值对的唯一类型化箭图为：





•属性图。

更一般地，给定一个多态映射 $\mathsf{s}: \mathbb{S} \to \mathbb{S}$，其上带有一个称为子类型关系的传递关系，则模式是 $\mathbb{S}$ 的一个模型Sch；而一个基于Sch 的类型化数据库由一个 $\mathbb{S}$ 的模型 G，以及 G 与 $U_{\mathsf{s}}(Sch)$ 之间在$Mod(\mathbb{S})$中的二元关系（称为类型关系）构成，并满足子类型公理：若 x 既是 x 的类型，又是 x' 的子类型，则 x' 是 x 的类型。于是，类型化数据库就是草图 $\mathbb{T}$ 的模型；该草图由 $\mathbb{S}$ 的一个副本、$\mathbb{S}'$ 的一个副本（其中 $\mathsf{s}_n: \mathbb{S} \to \mathbb{S}'$ 是 s 的分子）、两者之间的关系 τ（类型关系）以及子类型公理构成。

记号。 x 是 x 的一种类型这一事实记作 $x{:}\ x$，而 x 是 x' 的子类型这一事实记作 $x \sqsubseteq x'$。因此，子类型公理指出，如果 $x{:}\ x$ 且 $x \sqsubseteq x'$，那么 $x{:}\ x'$。

例如，从箭图出发，可以得到以类型化箭图为基础的草图导向数据库范式，并可通过添加属性-值对对其进行扩展。因此，我们提出以下将属性图定义为草图导向数据库范式的定义。

定义1。 草图导向属性图是一个在节点和边上带有属性-值对的类型化箭图。

3 面向图式的推理

各种图数据库的一个重要用途是，可以通过拼接连续边来构造路径。向图中添加所有路径通常会得到一个非常大的图，甚至是无限图，尽管通常只需要部分路径。例如，可以扩展箭图的图式 $\mathbb{S}_q$，使其包含用于构造所有路径的全拼接，或用于构造部分路径的部分拼接，而这两个图式之间由一个多态映射关联（详情见第 3.3 节）。与 s_{cq} 相关联的自由函子将拼接约束为全拼接，可用于按需生成路径。这是一个一般过程的实例：如下文所定义，局部化器是一个多态映射 $\mathsf{s}: \mathbb{R} \to \mathbb{S}$，它决定一个面向图式的推理过程。第 3.1 节回顾了一些范畴工具并定义局部化器；第 3.2 节说明其在数据库推理系统中的应用；第 3.3 节则将其应用于路径的构造。

3.1 图式的局部化器

在范畴论中，局部化是一个为某些箭头添加逆的函子。类似地，我们称图式态射在为某些箭头添加逆时为局部化。接着，我们将图式的局部化器 定义为“等价意义下的局部化”，具体而言，就是分子由局部化和等价态射复合而成的多态映射。由此可知，局部化器的复合仍是局部化器。设 $\mathsf{s}: \mathbb{R} \to \mathbb{S}$ 是一个局部化器，$F_{\mathsf{s}} \dashv U_{\mathsf{s}}$ 是其关联的伴随，则自由函子 F_{s} 是一个局部化。因此，根据伴随中局部化的一个重要属性，底函子 U_{s} 是忠实满的，并且 $F_{\mathsf{s}} \circ U_{\mathsf{s}}$ 是一个同构。

例如，我们来验证一个态射 $\mathsf{s}: \mathbb{R} \to \mathbb{S}$，它将 $\mathbb{R}$ 中的一个锥（无论是否交换）变为 $\mathbb{S}$ 中的一个极限，是局部化器。首先，如果 s 在 $\mathbb{S}$ 中为某些平行箭头 f和 g 添加了一个方程 $f \equiv g$，那么它就是局部化器，因为它使 f 和 g 的等化子箭头可逆。因此，如果 s 将 $\mathbb{R}$ 的任意有限锥变为交换锥，它就是局部化器。接着，如果 s 将一个交换锥的

$\mathbb{R}$ 变为一个极限，那么它就是局部化器，因为它使从给定锥的顶点到所构造极限的顶点的泛箭头可逆。

对于任意图式 $\mathbb{S}$，其底层箭图 $\mathbb{Q}$ 在 $\mathbb{S}$ 中的包含是一个局部化器。因此，计算 $Mod(\mathbb{S})$ 中图表 D 的余极限可以分两步进行：首先逐点计算D 在 $Mod(\mathbb{Q})$ 中的余极限，然后应用与 $\mathbb{Q}$ 在 $\mathbb{S}$中的包含相关联的自由函子。

图式的局部化器 $\mathsf{s}:\mathbb{R}\to\mathbb{S}$ 本质上会将 $\mathbb{R}$ 中的某些箭头变为 $\mathbb{S}$ 中的可逆箭头，这一事实可以借助“米田对应”用 $\mathbb{R}$的模型来表达。给定一个图式 $\mathbb{R}$， $\mathbb{R}$ 的反变米田模型是 $\mathbb{R}$ 在 $\mathbf{Mod}(\mathbb{R})$ 中的反变模型 $\mathbf{Yon}_{\mathbb{R}}$，满足对 $\mathbb{R}$中的每个点 X，$\mathbf{Yon}_{\mathbb{R}}(X)=\mathbf{Hom}^{Th(\mathbb{R})}(X,-)$；并且对 $\mathbb{R}$中的每个箭头f，$\mathbf{Yon}_{\mathbb{R}}(f)=-\circ f$。那么，对于每个多态映射 $\mathsf{s}:\mathbb{R}\to\mathbb{S}$，自由函子 F_{s} 满足 $F_{\mathsf{s}}\circ\mathbf{Yon}_{\mathbb{R}}$ 和 $\mathbf{Yon}_{\mathbb{S}}\circ\mathsf{s}$ 等价。米田引理断言，对于 $\mathbb{R}$ 中的每个点 X 以及$\mathbb{R}$的每个模型 M，在 X和 M中自然地存在一个双射，连接 $M(X)$ 与 $\mathbf{Hom}^{Mod(\mathbb{R})}(\mathbf{Yon}_{\mathbb{R}}(X)\ M)$。从 $\mathbf{Yon}_{\mathbb{R}}(X)$ 到 M 的模型态射称为 M中 X 的出现，因此米田引理建立了 M 中类型为 X 的元素与 M中 X 的出现之间的双射。

3.2 面向图式的推理系统

我们将面向图式的推理系统定义为一个局部化器 $\mathsf{s}:\mathbb{R}\to\mathbb{S}$；相对于 s 的面向图式的推理规则（分子为 s_n）则定义为 r :结论 $\to$ 前提中的一个箭头，其中 $\mathbb{R}$ 满足 $\mathsf{s}_n(r)$ 可逆。于是，前提和结论分别称为该规则r的前提和结论。 $\mathbb{R}$ 的一个模型 M 是模型 $F_{\mathsf{s}}(M)$ 的一个表示，其中模型取自 $\mathbb{S}$。推理过程旨在修改表示而不改变其所表示的模型。下文将此过程描述为范畴$Mod(\mathbb{R})$中的一个余极限。

给定一个相对于推理系统 $\mathsf{s}:\mathbb{R}\to\mathbb{S}$ 的规则 r : 结论 $\to$ 前提，它在 Yoneda 逆变模型下的像是一个态射 $\mathcal{Y}on_{\mathbb{R}}(r)$:$\mathcal{Y}on_{\mathbb{R}}$(前提) $\to\mathcal{Y}on_{\mathbb{R}}$(结论)，我们称之为与 r 相关联的语义规则。于是，给定 $\mathbb{R}$ 的一个模型 M，以及前提在M中的一个出现 p，将规则 r 应用于该出现 p，即在 $Mod(\mathbb{R})$ 中构造如下推出（其中 $\mathcal{Y}on=\mathcal{Y}on_{\mathbb{R}}$）：

$$\begin{array}{ccc}\mathcal{Y}on(Prem) & \xrightarrow{\mathcal{Y}on(r)} & \mathcal{Y}on(Conc)\\ {\scriptstyle p}\downarrow & & \downarrow{\scriptstyle q}\\ M & \xrightarrow{\quad m\quad} & M_1\end{array}$$

接下来可以验证， M_1与 M一样，是 $F_{\mathsf{s}}(M)$的一个表示。首先，态射$F_{\mathsf{s}}(\mathcal{Y}on(r))$是一个同构，因为它等于$\mathcal{Y}on_{\mathbb{S}}(F_{\mathsf{s}}(r))$，且 $F_{\mathsf{s}}(r)$是一个同构。因此， $F_{\mathsf{s}}(m)$也是一个同构，因为左伴随函子F_{s} 将推出映至$Mod(\mathbb{S})$中的一个推出。

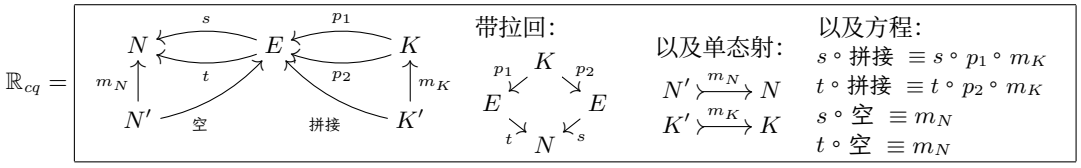
以图式为导向的推理规则与“通常的”逻辑规则之间的对应关系如下。 $\mathbb{R}$ 的模型就是该逻辑的公式，而以图式为导向的规则 r 对应于逻辑规则 Yon(Prem) Yon(Conc) 。现在，假设在 $\mathbf{Mod}(\mathbb{R})$ 中存在一种“良好”的并的概念（第 4 节将探讨这个问题）。给定一个逻辑规则 $\mathsf{P}_1\ldots\mathsf{P}_n\mathsf{C}_0$ (其中 $n\geq 0$)，令 $\mathsf{P}=\mathsf{P}_1\cup\cdots\cup\mathsf{P}_n$，并令 $\mathsf{C}=\mathsf{P}\cup\mathsf{C}_0$，其包含为 $\rho:\overline{\mathsf{P}\to\mathsf{C}}$。则相应的以图式为导向的规则是 $\mathbb{R}$ 的理论中的箭头 r，使得 $\rho=\mathbf{Yon}(r)$。

例如，面向图式的推理系统在图数据库中的一项重要应用是路径生成，见第 3.3 节。类型层次丰富的类型化数据库还有另一项重要应用，即定型过程。

3.3 路径

路径是图数据库特有的，是许多算法的基本组成部分，例如路径查找算法。本节将从面向图式的角度介绍如何在箭图上生成路径。这很容易扩展到标记箭图，但扩展到强标记箭图时则需要谨慎处理。这一点很重要，因为强标记箭图构成了资源描述框架（RDF）图（第 2.3 节）和 ER 图（第 2.3 节）的结构。

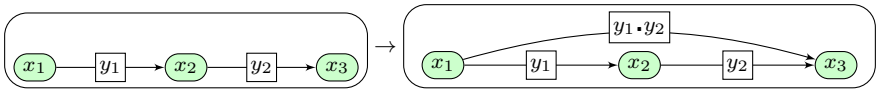
- **箭图上的路径**。
  在推理系统 $\mathsf{s}^{cq}:\mathbb{R}_{cq}\to\mathbb{S}_{cq}$ 中，草图 $\mathbb{R}_{cq}$ 为：



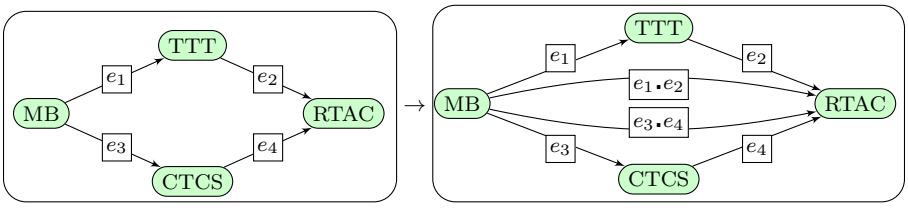
通过将 $N\prime$与 N识别，并将 K'与 K识别，令 m_N和 m_K分别对应于 id_N和 id_K，再添加方程以确保拼接 满足结合律，且以空 为单位元，我们便得到一个草图 $\mathbb{S}_{cq}$，它描述了包含所有路径的箭图（也就是说，一个小范畴的草图），以及一个作为局部化器的多态映射 $\mathsf{s}^{cq}:\mathbb{R}_{cq}\to\mathbb{S}_{cq}$。

记号。 为简洁起见，拼接也记作 $e_1.e_2=$拼接$(e_1,\ e_2)$。

例7。 与规则 $m_K:\ K'\to K$ 相关联的语义规则为：



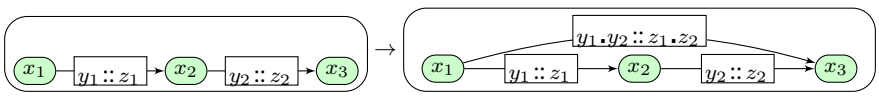
下方左侧的箭图中，该规则的前提有两处不同的出现。依次将规则应用于这两处，得到右侧的箭图：



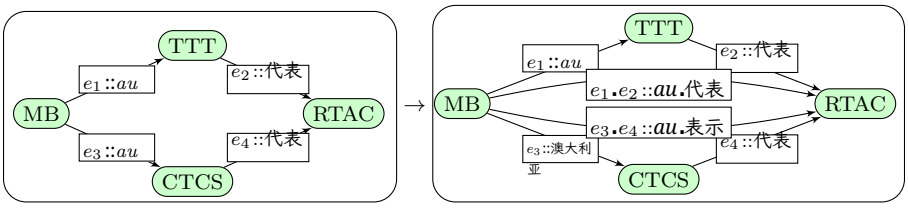
- **标记箭图上的路径**。
  假设所有标签都可以拼接，箭图上的路径很容易推广为标记箭图上的路径。因此，局部化器 s^{cq} 扩展为一个局部化器 $\mathsf{s}^{\ell cq}:\mathbb{R}_{\ell cq}\to\mathbb{S}_{\ell cq}$，它在边和标签上的拼接是相容的。

记号。 边的拼接记号很容易扩展到标记边：$e_1.e_2::\ell_1.\ell_2=$拼接$(e_1::\ell_1, e_2::\ell_2)$。

例8。 现在将例7扩展到标记箭图。与$m_K:K'\to K$ 相关联的语义规则是：



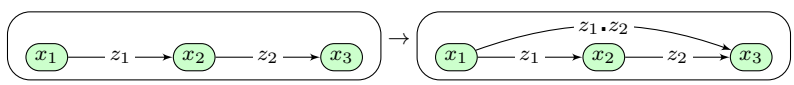
将此规则应用于其前提的两处出现后，我们得到：

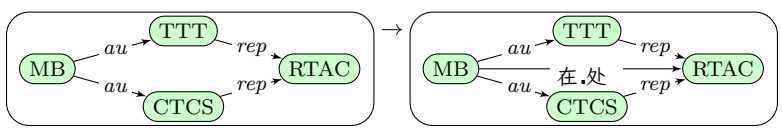


- **强标记箭图上的路径。**

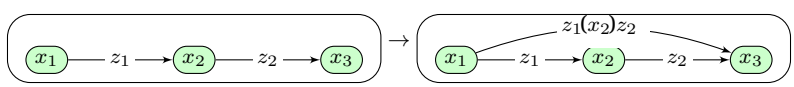
在强标记箭图中，有两种不同的方式来定义边的拼接，分别对应路径的两种不同预期含义。在第一种定义中，标签的拼接是标签上的二元运算，与标记箭图中的情况一样，记作$\ell_1 \centerdot \ell_2 =$ 拼接(ℓ_1, ℓ_2)。根据拼接的结合律，我们得到形如$(n_1) - \ell_1 \centerdot \ell_2 \centerdot \ldots \centerdot \ell_k \to (n_{k+1})$的路径。在第二种定义中，标签的拼接将两个标签 ℓ_1、ℓ_2 和一个节点 n 映射为一个记作 $\ell_1(n)\ell_2$的标签。根据拼接的结合律，我们得到形如$(n_1) - \ell_1(n_2)\ell_2 \ldots (n_k)\ell_k \to (n_{k+1})$的路径。

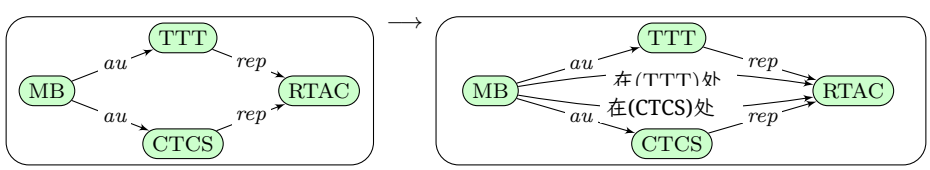
例9. 根据标签串接的第一个定义，下面依次给出了语义规则及其对该规则前提的两次出现的应用，两者都得出相同的结果。





根据标签串接的第二个定义，下面依次给出语义规则及其对前提两次出现的应用，由此得到两条路径。





- **属性图上的路径。**

将以图式为导向的属性图定义为一种类型化箭图，其节点和边上带有属性-值对。于是，类型构成一个强标记箭图（即模式），因此我们必须从上述两个定义中选择一个来定义标签的拼接。我们建议采用第二个定义，因为它似乎最为常见，这样路径的形式便为 $(n_1)\ell_1(n_2)\ell_2 \ldots (n_k)\ell_k(n_{k+1})$。

4 停顿图式

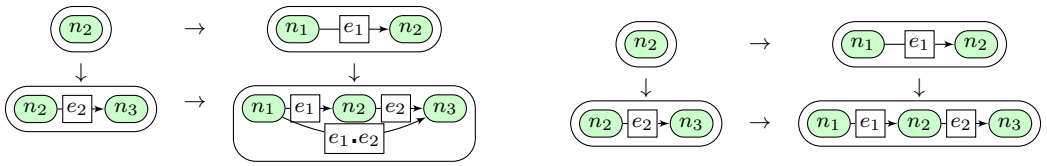
第2.1节中定义的图上的关系这一概念，在第2节和第3节中与数据库相关的许多图式中都有应用。它由两个嵌套的极限构成：先取图表上的极限，再取一个单态射。在集合论中，关系被定义为笛卡尔积的子集；众所周知，可以将关系定义为一个联合单态射函数族，从而省去笛卡尔积。本节探讨一个类似的问题：我们将停顿图式定义为有限极限图式，其中图表 D上的关系只通过一个极限锥来指定，从而避免使用 D的极限。一个显著的性质是，停顿图式的模型之并是逐点余极限。据我们所知，这为图式理论带来了一个新概念和一项新结果。在第4.1节中，我们定义图式模型的逐点极限与余极限。本文中的所有图式都是有限极限图式，因此模型的所有极限都是逐点的。然而，余极限并非如此；我们将验证，余极限具有逐点性这一性质在图式等价下并不稳定。在第4.2节中，我们将停顿图式定义为一类特定的有限极限图式，其中关系由一个极限而非两个嵌套的极限唯一指定（定理1）。随后，我们证明，停顿图式的模型的有限并是逐点余极限（定理 2）。

4.1 模型的逐点余极限

图式模型范畴中的逐点极限与余极限定义如下。设$\mathbb{S}_{set}$ 为仅由一个点 X_{set} 构成且不含任何箭头的图式，因此$Mod(\mathbb{S}_{set})$是集合范畴。对于每个图式 $\mathbb{S}$ 及其中的每个点 X，令 s_X：$\mathbb{S}_{set} \to \mathbb{S}$ 为将 X_{set} 映射到 X的态射。那么， $\mathbb{S}$ 模型的极限或余极限若在每个点 X 处都被与 s_X 相关联的底函子保留，就称为逐点 的。注意，这一属性取决于 $\mathbb{S}$的点，因此在图式等价下不保持不变：参见例10和例11。

图式模型范畴中的极限与余极限可用于定义集合论中包含、交与并等概念的类似概念。$\mathbb{S}$模型的包含 按逐点方式定义：若存在态射 $m\colon M\prime \to M$，使得对于 $\mathbb{S}$的每个点 X，函数 $m_X\colon M'(X) \to M(X)$都是包含映射，则模型 $M\prime$包含于模型 M中，记作 $M\prime \subseteq M$。此时 m 是$Mod(\mathbb{S})$中的单态射。给定 $\mathbb{S}$的一些模型，若它们都包含于某个公共模型 M_* 中，我们就称它们相容 。设 M_1和 M_2 是 $\mathbb{S}$的两个相容模型。那么，交 是包含映射 $M_1 \subseteq M_* \supseteq M_2$的拉回顶点。由于它是一个极限，因此它是逐点的。接着，并定义为 $Mod(\mathbb{S})$中包含映射 $M_1 \supseteq M_1 \cap M_2 \subseteq M_2$ 的推出。该推出存在，但通常不是逐点的。

例 10. 我们考虑用于从部分拼接到完全拼接的推理系统 $\mathsf{s}_{cq}\colon \mathbb{R}_{cq} \to \mathbb{S}_{cq}$，如第 3.3 节所示。以下交换方块展示了并的计算：先是在$Mod(\mathbb{S}_{cq})$中（左侧），然后是在$Mod(\mathbb{R}_{cq})$中（右侧）。为便于阅读，省略了空路径。



左侧的推出表示$Mod(\mathbb{S}_{cq})$中的并，显然不是逐点的，因为推出顶点处的边$[e_1 \,.\, e_2]$ 没有原像。右侧在$Mod(\mathbb{R}_{cq})$中的推出同样不是逐点的，尽管表面上看起来如此。请回想，该推出中的每个对象都解释了草图 $\mathbb{R}_{cq}$中的点N、 N'、 E、 K和 K'。特别是， K在推出顶点处被解释为 $K = \{(e_1、\ e_2)\}$，表示一对连续边（ e_1、 e_2）；而 K 在推出基中的每个对象处都被解释为空集。

4.2 停顿草图

记号。 以顶点 V 和投影 $p_X : V \to X$ 为图表 D 上的锥 C，记作 $p : V \to D$。图表 D 上关系的定义（见第 2.1 节）基于两个连续的极限：首先取 D 的极限，然后取该极限顶点上的一个单态射。本节将说明，只需一个极限即可定义它（定理1）。

以下关于停顿图表 和停顿锥的定义2和定义3，以及定理1，在具有所需极限的任意范畴中都成立。

定义2 （停顿图表）。 给定一个以 D 为基的交换锥 C，与 C 关联的停顿图表 记作 $C +_D C$，由 C 的两个副本沿着 D 粘合而成。

设 $C = \underline{p}\colon V \to D$，其中 D 记作 $X \ldots Y$，则：

$$C +_D C = \quad \begin{array}{ccccc} V & & & & V \\ {\scriptstyle p_X}\downarrow & & {\scriptstyle p_Y}\ \ {\scriptstyle p_X} & & \downarrow{\scriptstyle p_Y} \\ X & & \ldots & & Y \end{array}$$

注记 2。 在 $C +_D C$ 上的任意交换锥 C' 都由其顶点 W 及投影 $q_1, q_2 \colon W \to V$ 唯一确定，这两个投影分别对应 V 的两个副本，并且必须满足：对于 D 中的每个 X，均有 $p_X \circ q_1 = p_X \circ q_2$。于是，对于 D 中的每个点 X，在 C' 中的投影 $q_X \colon W \to X$ 为 $q_X = p_X \circ q_1 = p_X \circ q_2$。

$$C' = \quad \begin{array}{ccccc} & & W & & \\ & {\scriptstyle q_1}\swarrow & & \searrow{\scriptstyle q_2} & \\ V & & {\scriptstyle p_Y}\ \ {\scriptstyle p_X} & & V \\ {\scriptstyle p_X}\downarrow & & & & \downarrow{\scriptstyle p_Y} \\ X & & \ldots & & Y \end{array}$$

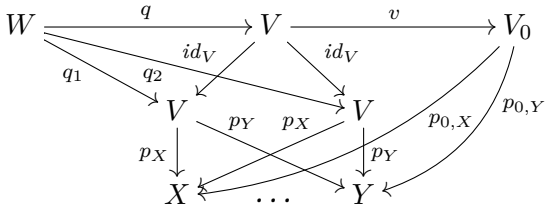
定义3 （停顿锥）。 给定以 D 为基的交换锥 $C =(p : V \to D)$，与 C 相关联的停顿锥记作$Stt(C)$，它是一个以 $C +_D C$ 为基、以 V为顶点的交换锥，其在 V 的两个副本上的投影均为$id_V \colon V \to V$，其他投影则为 $p_X \colon V \to X$，与 C 相同。以 D 为基的停顿锥是指与某个以 D 为基的交换锥相关联的停顿锥。

$$Stt(C) = \quad \begin{array}{ccccc} & & V & & \\ & {\scriptstyle id_V}\swarrow & & \searrow{\scriptstyle id_V} & \\ V & & {\scriptstyle p_Y}\ \ {\scriptstyle p_X} & & V \\ {\scriptstyle p_X}\downarrow & & & & \downarrow{\scriptstyle p_Y} \\ X & & \ldots & & Y \end{array}$$

在集合范畴中，定理1意味着，通常定义为笛卡尔积（涉及两个嵌套极限）的子集的关系，也可以定义为唯一极限的顶点。

定理1. 设 D 为一个图表， $C_0 = (p_0 : V_0 \to D)$ 为 D 的极限，且 $C = (p : V \to D)$ 为 D上的任意交换锥。设 $v : V \to V_0$ 为唯一的箭头，使得对 D的每个点 X，有 $p_X = p_{0,X} \circ v : V \to X$。则 $\overline{v}$ 是单态射，当且仅当 $Stt(C)$ 是 $C +_D C$的极限。

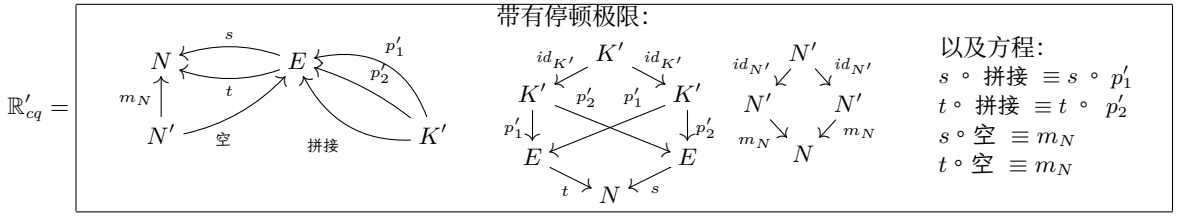
证明。利用备注 2，Stt(C) 是 $C +_D C$ 的极限，当且仅当对于图表$C +_D C$上的每个锥，其顶点为 W，投影为 $q_1, q_2 \colon W \to V$，存在唯一的态射 $q \colon W \to V$，使得 $q_1 = q$ 且 $q_2 = q$。等价地，当且仅当对于每个 $q_1, q_2 : W \to V$，若对 D 中的每个 X 有$p_X \circ q_1 = p_X \circ q_2$，则有 $q_1 = q_2$。条件 $p_X \circ q_1 = p_X \circ q_2$ 可写为 $p_{0,X} \circ v \circ q_1 = p_{0,X} \circ v \circ q_2$。由于 C_0 是 D 的极限，该条件对 D 中的每个 X成立，当且仅当 $v \circ q_1 = v \circ q_2$。因此，Stt($C$) 是 $C +_D C$ 的极限，当且仅当对于每个q_1, $q_2 \colon W \to V$，若 $v \circ q_1 = v \circ q_2$，则有 $q_1 = q_2$：这意味着 v 是单态射。



□

定义 4（停顿图式）。停顿图式是指其所有潜在极限都是停顿锥的草图。

例如，对于带部分串接的箭图，下方的停顿图式 $\mathbb{R}'_{cq}$ 与 $\mathbb{R}_{cq}$等价，其定义见第 3.3 节。



定理 2. 停顿图式的相容模型的有限并是逐点余极限。

证明。只需对两个模型证明这一结果。在范畴 Mod(𝕊) 中，交$M_1 \cap M_2$ 是逐点的，而并 $M_1 \cup M_2$ 是包含映射 $M_1 \supseteq M_1 \cap M_2 \subseteq M_2$ 的余极限，通常并非逐点的。令 $\mathbb{Q}$ 为 $\mathbb{S}$ 的底层箭图，且 $\mathsf{s} : \mathbb{Q} \to \mathbb{S}$ 为包含映射。于是，M_1 与 M_2 的交在 Mod(𝕊) 和 Mod($\mathbb{Q}$) 中相同，而 Mod(𝕊) 中 $M_1 \supseteq M_1 \cap M_2 \subseteq M_2$ 的余极限 M_3 是 Mod($\mathbb{Q}$) 中 $M_1 \supseteq M_1 \cap M_2 \subseteq M_2$ 的余极限 N_3 在 F_{s} 下的像。因此，$N_3 = M_1 \cup M_2$ 在 Mod($\mathbb{Q}$) 中；若 N_3 在 Mod(𝕊) 中，则 $M_3 = N_3 = M_1 \cup M_2$ 在 Mod(𝕊) 中。现在证明 N_3 在 Mod(𝕊) 中，这意味着它将 $\mathbb{S}$ 的每个极限映为集合范畴中的一个极限。考虑 $\mathbb{S}$ 的任意潜在极限，它是一个锥 $C' =$Stt(C)，其基为 $C +_D C$；这里的图表为 D，而锥 C 的基为 D。令 V 为 C 和 C' 的公共顶点。令 $\mathbb{S}'$ 由 $\mathbb{S}$ 添加 D 的一个极限得到，其顶点记为 V_0，并添加相应的典范箭头 $v : V \to V_0$ 及相应方程。于是，$\mathbb{S}'$ 等价于 $\mathbb{S}$。根据定理1，我们只需证明 $N_3(v)$ 是单射。我们有 $N_3(V) = M_1(V) \cup M_2(V)$，其中 $M_1(V) \subseteq M_1(V_0)$ 且 $M_2(V) \subseteq M_2(V_0)$，因此 $N_3(V) \subseteq M_1(V_0) \cup M_2(V_0)$。由于 $M_1(V_0) \cup M_2(V_0) = N_3(V_0)$，可得 $N_3(V) \subseteq N_3(V_0)$。因此，$N_3$ 是 $\mathbb{S}$ 的一个模型，从而 $M_3 = N_3$ 和 M_3 是 Mod(𝕊) 中的逐点余极限，证毕。

□

例 11. 在例 10 的右侧方形中，$Mod(\mathbb{R}_{cq})$ 中计算出的并不是逐点并，因为点 K 来自 $\mathbb{R}_{cq}$的解释如此。不过，由于 $\mathbb{R}'_{cq}$ 等价于 $\mathbb{R}_{cq}$，这个右侧方形也是 $Mod(\mathbb{R}'_{cq})$ 中的一个并，而且这个并是逐点的。

5 结论

本文展示了以有限极限草图为基础构建的以图式为导向的数据库，如何为图数据库建模提供强大而灵活的框架。借助范畴论，我们证明了有限极限草图能够有效表示各种数据库范式，包括 RDF 图和属性图。通过对关系和表进行编码，可以轻松将其应用于关系数据库（见附录 A）。草图的图示特性与既有的数据库形式体系十分接近，例如实体关系图、UML 和图形化模式语言。我们还探讨了推理系统、路径、模式和类型化等高级特性，凸显了它们与以图式为导向的方法的兼容性。停顿草图的引入提供了一种简化数据库结构并确保模型逐点并的新方法，有助于应对复杂数据关系管理中的挑战。这一概念为指定数据库操作以及维护数据库间的一致性开辟了新的途径。面向草图的数据库将理论严谨性与实际适用性融为一体，使其成为表示和查询复杂数据结构的有前景工具。未来的工作包括将所提出的框架扩展到代数查询应答，以及进一步研究其在语义网和基于本体的系统中的应用。

参考资料

[1] Renzo Angles。《属性图数据库模型》。*Alberto Mendelzon* 数据管理基础国际研讨会（*AMW*），2018 年。以综述形式介绍属性图模型、其变体和查询模式；有助于对比资源描述框架（RDF）三元组存储与属性图。

[2]Kenneth Baclawski、Dan A. Simovici 和 William White。《数据库语义的范畴方法》。《计算机科学中的数学结构》，4(2):147–183，1994 年。URL：https://www.cambridge.org/core/journals/mathematical-structures-in-computer-science/article/abs/categorical-approach-to-database-semantics/247498068CAD4929C5378E91DE5D34F1、doi:10.1017/S0960129500000426。

[3] Michael Barr。《Horn 理论的模型》。载于 J. W. Gray 编，《style id='2'>《计算机科学与逻辑中的范畴》，《当代数学》第 92 卷，第 1–7 页。美国数学会，普罗维登斯，罗得岛州，1989 年。已在各种文集中转载或重新发行；可在线免费获取。URL: https://ncatlab.org/nlab/files/Barr-ModelsOfHornTheories.pdf。

[4] Michael Barr 和 Charles Wells。《托扑斯、三元组与理论》，《数学科学基础丛书》第 278 卷。施普林格出版社，纽约／柏林，1985 年。转载于《范畴论的理论与应用重印本》第 12 期（2005 年）。网址： https://www.math.mcgill.ca/barr/papers/tttall.pdf。

[5] Michael Barr 和 Charles Wells。《计算机科学中的范畴论》。普伦蒂蒂·霍尔，英国海默尔亨普斯特德，1990 年。转载于《范畴论的理论与应用重印本》第 22 期（2012 年）；可在线免费获取。网址： http://www.tac.mta.ca/tac/reprints/articles/22/tr22.pdf。

[6]Francis Borceux.《范畴代数手册》，《数学及其应用百科全书》第50卷。剑桥大学出版社，剑桥，重印版，

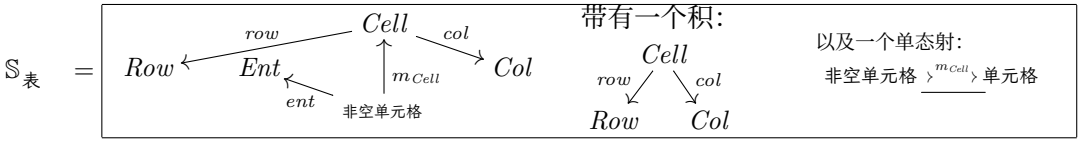
1994年。共3卷：第1卷《基础范畴论》；第2卷《范畴与结构》；第3卷《层范畴》。网址：https://www.cambridge.org/core/series/encyclopaedia-of-mathematics-and-its-applications/50.[7] Boris Cadish 和 Zinovy Diskin. 基于图式和方程的异构视图集成。载于《智能系统基础：第十届国际研讨会 *ISMIS 1996* 论文集》，《计算机科学讲义》第1079卷，第603–612页。施普林格，1996年。[8] E. F. Codd. 大型共享数据库的数据关系模型。《美国计算机协会通讯》，13(6):377–387，1970年。这篇奠基性论文提出了数据关系模型，是理解基于表的模式和基于 SQL 的系统的关键。doi:10.1145/362384.362685.[9] C. J. Date.《数据库系统导论》。Addison-Wesley，第8版，2003年。这本全面的教科书介绍了关系理论、规范化、查询和数据库设计实践，可作为权威教材参考。[10] Irena Holubov´a、Pavel Koupil 和 Martin Svoboda. 多模型数据的范畴建模：一统天下的模型。载于第十届模型与数据工程国际会议（*MEDI 2021*）论文集，爱沙尼亚塔林，2021年。[11] ISO/IEC. GQL——图查询语言（ISO/IEC 39075），2024年。近期一项国际标准化工作，旨在制定一种图查询语言，使其可与关系系统中的 SQL 相媲美。网址：https://www.iso.org/standard/76120.html.[12] Michael Johnson 和 Robert Rosebrugh. 图式数据模型、关系模式与数据规格。载于*CATS'02* 论文集：计算：澳大拉西亚理论研讨会，《理论计算机科学电子札记》第61卷，第51–63页，荷兰阿姆斯特丹，2002年。爱思唯尔。访问日期：2025-10-13。网址：https://mta.ca/~rrosebru/articles/sdmrsds.pdf.[13] Michael Johnson、Robert Rosebrugh 和 R. J. Wood. 实体-关系-属性设计与图式。《范畴理论及其应用》，10(3):94–112，2002年。EA 图式；亦可见 http://eudml.org/doc/121924。网址：http://www.tac.mta.ca/tac/volumes/10/3/10-03/10-03.pdf.[14] Saunders Mac Lane.《工作数学家的范畴论》，《数学研究生教材》第5卷。施普林格，纽约，第2版，1998年。网址：https://link.springer.com/book/10.1007/978-1-4757-4721-8, doi:10.1007/978-1-4757-4721-8.[15] S. K. Lellahi 和 Nicolas Spyratos. 支持结构化对象和继承的范畴数据模型。载于新一代信息系统技术：东西方数据库研讨会（*EWDW*），《计算机科学讲义》第504卷，第86–105页。施普林格，1990年。[16] Zhen Hua Liu、Jiaheng Lu、Dieter Gawlick、Heli Helskyaho、Gregory Pogossiants 和 Zhe Wu. 多模型数据库管理系统——展望未来。载于异构数据管理、多存储系统与医疗保健分析（*DMAH 2018，Poly 2018*），《计算机科学讲义》第11470卷，第16–29页。施普林格，2019年。[17] Jiaheng Lu、Zhen Hua Liu、Heli Helskyaho、Jiahua Xu 和 Yannis Velegrakis. 多模型数据库：应对数据多样性的新征程。《*SIGMOD* 记录》，47(2):23–33，2018年。doi:10.1145/3299887.3299891。

[18] Zoran Majki´c 和 Bhanu Prasad。数据库映射的图式理论。《智能信息与数据库系统国际期刊》，9(1):32–53，2015。URL: https://www.inderscience.com/info/inarticle.php?artid=70807， doi:10.1504/IJIIDS.2015.070807。[19] Robert Rosebrugh 和 R. J. Wood。关系数据库与指标范畴。载于计算机科学范畴方法研讨会，加拿大蒙特利尔，1991。URL: https://mta.ca/~rrosebru/articles/rdic.pdf。[20] David I. Spivak 和 Ryan Wisnesky。函子式数据迁移的关系基础。载于 James Cheney 和 Thomas Neumann 编，第 *15* 届数据库编程语言研讨会论文集，美国宾夕法尼亚州匹兹堡，*2015* 年 *10* 月 *25–30* 日，第 21–28 页。ACM，2015。doi:10.1145/2815072.2815075。[21] Arthur H. M. ter Hofstede、E. Lippe 和 Paul J. M. Frederiks。从范畴论视角进行概念数据建模。《计算机杂志》，39(3):215–231，1996。URL: https://academic.oup.com/comjnl/article/39/3/215/399043， doi:10.1093/comjnl/39.3.215。[22] Laurent Thiry、Heng Zhao 和 Michel Hassenforder。（大）数据模型的范畴与优化。《大数据期刊》，5(21)，2018。URL: https://journalofbigdata.springeropen.com/articles/10.1186/s40537-018-0132-9， doi:10.1186/s40537-018-0132-9。[23] Chris Tuijn 和 Marc Gyssens。从范畴论视角看数据库的视图与分解。载于数据库理论——*ICDT'92*，第四届国际会议，德国柏林，*1992* 年 *10* 月 *14–16* 日，论文集，《style id='32'>计算机科学讲义第 646 卷，第 99–112 页。施普林格，1992。[24] Chris Tuijn 和 Marc Gyssens。Cgood：一种范畴图导向对象数据模型。《理论计算机科学》，160(1-2):217–239，1996。doi:10.1016/0304-3975(95)00089-5。[25] W3C。SPARQL 1.1 查询语言，2013。用于 RDF 图的标准化查询语言；对语义网查询和推理示例十分重要。URL: https://www.w3.org/TR/sparql11-query/。[26] W3C。资源描述框架（RDF）1.1 概念与抽象句法，2014。W3C 规范及对 RDF 三元组和基于图的语义数据建模的概念性介绍。URL: https://www.w3.org/TR/rdf11-concepts/。[27] Ryan Wisnesky、Spencer Breiner、Albert T. Jones、David I. Spivak 和 Eswaran Subrahmanian。利用范畴论促进多个制造服务数据库集成。《计算机、信息与工程科学杂志》，17(2)，2017。URL: https://doi.org/10.1115/1.4034268。

一个关系数据库

尽管本文侧重于图数据库，但值得注意的是，关系数据库范式也可以视为以图式为导向的。在本附录中，我们提出一些可用于此目的的提示。还可以加入更多问题，例如添加涉及多个表的主键和外键约束。

一个表由一组行、一组列和一组条目组成，每行 x 和每列 y至多有一个条目$e(x,\ y)$。因此，表是以下草图的一个模型，其中行、列、实体、单元格和非空单元格分别解释为行、列、实体、单元格和非空单元格的集合。

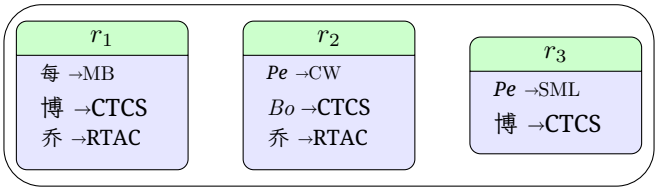


此外，可以观察到，每个表都可以关联到一个节点上带有属性-值对的箭图，如下所示：没有边；表的每一行对应一个节点；每一列对应一个属性；每个条目对应一个值；对于每个非空单元格（r, c），节点 r 上都有一个属性-值对 $c \to v$，其中 $v = \text{entry}(r, c)$。因此，与第 2.3 节一样，每个表都可以关联到一个强标记箭图，其中表的每一行对应一个节点，每个条目也对应一个节点；每一列对应一个标签；对于每个非空单元格（r, c），都有一条边（r）$-c\to$（v），其中 $v = \text{entry}(r, c)$。每个表都关联到一个节点上带有属性-值对的箭图这一事实，可以看作是将与从$\mathbb{S}_{av},_X$ 到 $\mathbb{S}_{table}$ 的多态映射相关联的底函子应用于该映射的结果；该映射将 X, V, A 映射到行、实体、列，因此将 T'映射到积 Row ×Ent ×Col（同构于 Cell ×Ent），并将 T 映射到非空单元格，将 m_T 映射到由 m_{Cell} 和 ent 构成的有序对。

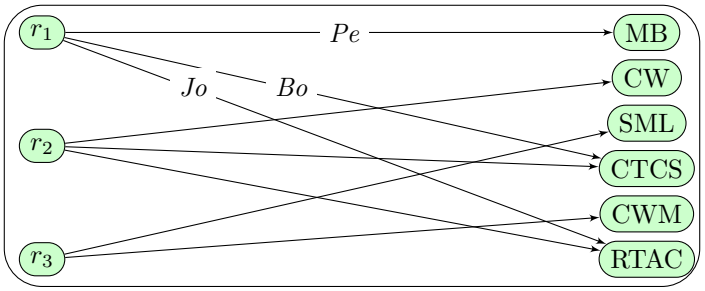
例 12. 例5中的部分数据现在以表的形式存储:

	Pe	*Bo*	*Jo*
r_1	MB	CTCS	RTAC
r_2	CW	CTCS	RTAC
r_3	SML	CWM	

以下是带有属性-值对的相关箭图:



和 相关的强标记箭图（为便于阅读，仅标注了部分标签） :



这种观点与第 2.3 节中的模式与类型赋予概念相符。

例 13. 例 12 的模式是这个单行表：

	Pe	Bo	Jo
Biblio	字符串	字符串	字符串

下面是带有属性-值对的关联箭图，以及关联的强标记箭图：

