反事实解释的公理基础

Leila Amgoud[a,*] **和 Martin Cooper**[b]

[a]IRIT，CNRS，法国[b]IRIT，图
卢兹大学，法国

arXiv:2602.04028v1 [cs.AI] 3 Feb 2026

摘要。对自主智能系统进行解释，对于提升人们对其决策的信任至关重要。反事实已成为最有说服力的解释形式之一。它们通过揭示决策可以如何改变，回答“为什么不”这一问题。尽管相关文献日益增多，现有大多数解释器只关注单一的反事实类型，且局限于局部解释，聚焦于个体实例。目前尚无对其他反事实类型或能够揭示系统整体推理过程的全局反事实进行系统研究。

本文通过引入一个以反事实解释器的一组理想属性为基础的公理化框架，弥补了这两方面的不足。文中证明了不可能性定理，表明不存在任何单一解释器能够同时满足某些公理组合，并完整刻画了所有相容的公理集合。随后，表示定理确立了特定公理子集与满足这些公理的解释器族之间的五组一一对应关系。每个解释器族都会产生一种不同类型的反事实解释，由此揭示出五种本质上不同的反事实类型。其中一些对应局部解释，另一些则体现全局解释。最后，该框架将现有解释器纳入这一分类体系，形式化刻画其行为，并分析生成此类解释的计算复杂度。

1 引言

解释自主智能系统的决策在过去十年中引起了广泛关注（参见 [34, 35] 中的综述）。本研究的总体目标是增强对此类系统所做决策的信任。两类主要的解释已得到广泛研究。第一类通过识别导致特定决策的关键因素来回答“为什么”问题（例如，[10, 13, 33]）。例如：“Alice没有资格获得贷款，因为她的年收入为 $20K。”第二类——反事实——通过描述如何改变该决策来回答“为什么不”问题。例如：“如果Alice的年收入为$35K，她就有资格获得贷款。”

Byrne在 [6, 7] 中指出，受决策影响的个人往往更偏好那些说明他们如何能够获得更好结果的解释。这一见解推动了大量反事实解释器的发展（参见 [24, 15, 22] 中近期的全面综述），这些解释器既适用于自主智能体，也适用于多智能体系统（例如，[38, 37, 9, 27]）。

绝大多数现有研究都聚焦于局部反事实——针对单个输入实例量身定制的解释——并且通常只考虑**单一类型的反事实**。对于给定输入，反事实被定义为能够产生不同结果的“最接近”替代输入。不同的解释器在如何形式化接近度这一概念上各不相同。为指导其定义以及“合理”反事实的选择，文献中提出了若干性质（例如，[5, 17, 44, 49, 47]）。例如，**最小性** 要求反事实对原始输入所做的改变尽可能小，而**多样性**则鼓励针对同一输入生成彼此不同的反事实（其他性质请参见第7节）。现有解释器通常根据其满足这些性质的程度，以及用于生成反事实的策略来评估。

尽管关于反事实解释的研究成果丰富，但尚未有人系统研究局部反事实的其他类型。此外，全局反事实旨在捕捉并解释模型整体的决策行为，却受到的关注少得多；迄今为止，唯一值得一提的贡献来自 [2]。因此，人们对局部反事实与全局反事实之间的理论权衡仍缺乏深入理解。

本文提出一个形式化框架，以系统地定义和比较不同类型的局部反事实与全局反事实，从而弥补上述不足。我们采用公理化方法——这是现代理论科学中一种成熟的方法，通过明确陈述公理（基本性质）来构建体系，并据此推导表示定理。将这种方法应用于反事实有多项优势。首先，它通过精确定义理想属性，带来**清晰性** 与**严谨性**，促进人们对何为可靠反事实形成共同认识。其次，它能够依据反事实类型所满足的公理，对其进行**系统比较**，从而明确呈现它们的差异、优势与局限。这有助于开展有原则且客观的评估，而不再仅仅依赖经验表现。最后，该框架**与模型无关**且**具有普适性**，确保研究结果广泛适用于各种情境。

本文的贡献有六个方面，可补充现有关于反事实解释的文献：

1. 本文提出了一种公理化框架，以九条公理为基础。2. 本文给出了不可能性定理，表明没有任何解释器能够同时满足某些特定的公理组合。3. 本文完整刻画了所有相容的公理集合。4. 本文提出了表示定理，定义了特定公理子集与满足这些公理的反事实解释器族之间的一一对应关系。这些刻画很有价值，因为它们精确且**全面**地

* 通讯作者。电子邮件：leila.amgoud@irit.fr

描述了**所有**满足给定公理集合的解释器。主要发现包括：

- 这九条公理区分出五种不同类型的反事实，每种反事实都由不同的解释器族生成。
- 反事实解释有两种基本形式：必要理由和充分理由。
- 对于每种形式（必要、充分），其中一种类型对应于全局解释，其余类型则侧重于局部解释。
- 在局部充分理由中，有两种子类型：怀疑型和确信型。

5. 该框架对现有的反事实解释器进行了形式化刻画，表明几乎所有解释器都属于生成局部确信式充分理由的解释器族。在 [2] 中提出的全局解释器对应于全局必要理由族。因此，该框架揭示了三种此前未被识别的反事实类型。

6. 该框架分析了生成解释的计算复杂度。

本文组织如下。第2节介绍必要的背景，第3节定义公理，第4节和第5节专门讨论表示定理。第6节比较各种类型，第7节讨论相关工作，第8节给出复杂度结果。本文所有命题和定理的证明均见附录。

2 背景

本文始终将分类理论视为一个元组$\mathbb{T} = \langle \mathrm{F}, \mathrm{dom}, \mathrm{C} \rangle$，它由有限特征集 F、函数 dom（返回每个特征 $f \in \mathrm{F}$ 的定义域，且 $\mathrm{dom}(f)$ 为有限集，并且 $|\mathrm{dom}(f)| > 1$）以及有限类别集 C（其中 $|\mathrm{C}| \geq 2$）组成。令 $\mathrm{Lit}(\mathbb{T})$ 表示 $\mathbb{T}$ 中所有文字的集合，其中一个文字是一个二元组 (f, v)，且 $f \in \mathrm{F}$、$v \in \mathrm{dom}(f)$。部分赋值是任意一个文字集合，其中每个特征在 F 中至多出现一次；若每个特征都恰好出现一次，则称其为实例。我们用 $\mathbb{E}(\mathbb{T})$ 表示所有可能的部分赋值的集合，用 $\mathbb{F}(\mathbb{T})$ 表示理论 $\mathbb{T}$的特征空间，即所有实例的集合。对于集合 $E \in \mathbb{E}(\mathbb{T})$，$\mathbb{E}(\mathbb{T}) \oslash E$ 表示不包含来自 E 的任何文字的所有部分赋值的集合；也就是说，对于任意 $X \in \mathbb{E}(\mathbb{T}) \oslash E$，$X \in \mathbb{E}(\mathbb{T})$ 且 $E \cap X = \emptyset$。对于 $E \in \mathbb{E}(\mathbb{T})$ 和 $x \in \mathbb{F}(\mathbb{T})$，我们用

$x_{\downarrow E}$ $x_{\downarrow E} = E \cup \{ f\ v \in x \mid$ 表示满足 $\nexists v' \in f\ v' \neq v\ f\ v' \in E\}$ (,) dom( ) s.t. and (,) . 的部分赋值。其思路是将共有特征的取值替换为 E 中的取值。考虑一个由两个二元特征组成的理论，实例 $x = ((f_1, 0), (f_2, 0))$ 和 $E = \{(f_2, 1)\}$。于是，$x_{\downarrow E} = ((f_1, 0), (f_2, 1))$。

性质 1. 令 $\mathbb{T}$ 为分类理论。$\forall x \in \mathbb{F}(\mathbb{T})$，$\forall E \in \mathbb{E}(\mathbb{T})$，$x_{\downarrow E} \in \mathbb{F}(\mathbb{T})$。

对于任意 $x \in \mathbb{F}\ \mathbb{T}$ $E \in \mathbb{E}\ \mathbb{T}$ $x \ominus E = \{y \in \mathbb{F}\ \mathbb{T} \mid ()$ 以及任意 ( )，( )。考虑一个由两个特征组成的理论，其中 $\mathrm{dom}(f_1) = \{0, 1\}$ 且 $\mathrm{dom}(f_2) = \{0, 1, 2\}$，实例 $x = ((f_1, 0), (f_2, 0))$ 和 $E = \{(f_2, 0)\}$。于是，$x \ominus E = \{y, z\}$ 其中 $y = ((f_1, 0), (f_2, 1))$ 和 $z = ((f_1, 0), (f_2, 2))$。

理论 $\mathbb{T}$ 上的分类器 是一个**满射** 函数 κ：$\mathbb{F}(\mathbb{T}) \to \mathrm{C}$，它将每个实例映射到一个类别，使得每个类别至少分配给一个实例。

在文字中，我们区分出在给定分类器下对于给定类别而言属于核心或必需的文字。

定义 1. 设 $\mathbb{T} = \langle \mathrm{F}, \mathrm{dom}, \mathrm{C} \rangle$ 是一个理论，$c \in \mathrm{C}$，κ 是定义在 $\mathbb{T}, l \in \mathrm{Lit}(\mathbb{T})$上的一个分类器。$l$ 是 c 的核心，当且仅当 $\forall x \in \mathbb{F}(\mathbb{T})$ 使得 $\kappa(x) = c, l \in x$。核心(c) 是 c的核心文字集合。

Core(c) 集合对于 $c \in \mathrm{C}$ 可能为空，如下所示。

例1。令 κ 为一个预测带朋友去哪里的分类器。它是温度 t和朋友最喜欢的活动 $\{ \quad \} = \{ \quad \}$ κ 的函数，其中 $\mathrm{dom}(t) = \{$炎热、温和、冰冻$\}$，dom() 攀岩、阅读、滑雪，以及 C 海滩、山、电影院。该分类器在炎热的日子预测去海滩；如果天气温和且朋友喜欢攀岩，或天气寒冷且她喜欢滑雪，则预测去山上；否则预测去电影院。下方列出了所有实例及其决策。

$\mathbb{F}(\mathbb{T})$	t	a	$\kappa(x_i)$
x_1	*hot*	攀岩	海滩
x_2	*mild*	攀岩	山
x_3	冰冻	阅读	电影院
x_4	冰冻	滑雪	山
x_5	冰冻	攀岩	电影院
x_6	*hot*	阅读	海滩
x_7	*hot*	滑雪	海滩
x_8	*mild*	阅读	电影院
x_9	*mild*	滑雪	电影院

以下列出了对应于每个类别的核心文字集合。

- 核心(海滩) $= \{(t,$ 炎热$)\}$,
- 核心(山) $= \emptyset$,
- 核心(电影院) $= \emptyset$.

设 X 为一个对象集合。在 X 上的预序 是 X 上的二元关系$\succeq$，它具有自反性 和传递性。当且仅当对于所有 $x, y \in X$，$x \succeq y$ $y \succeq x$ $x \succeq y$ x 或 时，它是全序 。记号 表示至少与 y 一样受偏好；$x \succ y$ 是 $x \succeq y$ 和 $y \not\succeq x$ 的简写。令 $\max(X, \succeq) = \{E \in X \mid \nexists E' \in X$，使得 $E' \succ E\}$ 成为 X 中最受偏好的元素集合。最后，集合 X 上的权重函数 是一个函数 σ：$X \to [0, +\infty)$。

3 公理化框架

我们提出一个公理化框架，用于定义反事实解释器–生成反事实解释的函数。在介绍公理之前，我们先引入查询这一概念，它表示向解释器提出的问题。

定义 2. 查询 是一个元组$\mathbf{Q} = \langle \mathbb{T}, \kappa, x \rangle$，满足 $\mathbb{T}$ 是一个分类理论，κ 是 $\mathbb{T}$ 上的分类器，并且 $x \in \mathbb{F}(\mathbb{T})$。

反事实解释器（简称解释器）是一个函数，它将每个查询映射为一组部分赋值，每个部分赋值都表示一个反事实。正如我们将在后续章节中看到的，反事实并没有唯一的定义。尽管如此，文献普遍认为，所有这类定义都旨在深入揭示如何改变分类器（或人工智能系统）所作出的预测。

定义 3. 反事实解释器 是一个函数$\mathbf{L}$，它将每个查询$\mathbf{Q} = \langle \mathbb{T}, \kappa, x \rangle$ 映射为 $\mathbb{E}(\mathbb{T})$的一个子集。每个 $E \in \mathbf{L}(\mathbf{Q})$ 都是一个反事实。

记号：设$\mathbf{L}$、$\mathbf{L}'$ 是两个解释器。我们用$\mathbf{L} \sqsubseteq \mathbf{L}'$ 表示：对于任意查询$\mathbf{Q}$，$\mathbf{L}(\mathbf{Q}) \subseteq \mathbf{L}'(\mathbf{Q})$。

General	成功性	$\forall \mathbf{Q} = \langle \mathbb{T}, \kappa, x \rangle$, $\mathbf{L}(\mathbf{Q}) \neq \emptyset$。
	非平凡性	$\forall \mathbf{Q} = \langle \mathbb{T}, \kappa, x \rangle$, $\emptyset \notin \mathbf{L}(\mathbf{Q})$。
	等价性	$\forall \mathbf{Q} = \langle \mathbb{T}, \kappa, x \rangle$, $\forall \mathbf{Q}' = \langle \mathbb{T}, \kappa, x' \rangle$, 如果 $\kappa(x) = \kappa(x')$，那么 $\mathbf{L}(\mathbf{Q}) = \mathbf{L}(\mathbf{Q}')$。
Nec. Reasons	可行性	$\forall$**问答** $= \langle \mathbb{T}, \kappa, x \rangle, \forall E \in \mathbf{L}($**问答**$), E \subseteq x.$
	核心性	$\forall \mathbf{Q} = \langle \mathbb{T}, \kappa, x \rangle, \forall E \in \mathbf{L}(\mathbf{Q}),\ E \subseteq \mathtt{Core}(\kappa(x))$.
	怀疑有效性	$\forall \mathbf{Q} = \langle \mathbb{T}, \kappa, x \rangle, \forall E \in \mathbf{L}(\mathbf{Q}),\ \forall y \in x \ominus E, \kappa(y) \neq \kappa(x)$.
Suff. Reasons	新颖性	$\forall \mathbf{Q} = \langle \mathbb{T}, \kappa, x \rangle$, $\forall E \in \mathbf{L}(\mathbf{Q})$, $x \cap E = \emptyset$。
	强有效性	$\forall \mathbf{Q} = \langle \mathbb{T}, \kappa, x \rangle$, $\forall E \in \mathbf{L}(\mathbf{Q})$, $\nexists y \in \mathbb{F}(\mathbb{T})$，满足 $E \subseteq y$ 且 $\kappa(y) = \kappa(x)$。
	弱有效性	$\forall \mathbf{Q} = \langle \mathbb{T}, \kappa, x \rangle$, $\forall E \in \mathbf{L}(\mathbf{Q})$, $\kappa(^{x}\!\downarrow_{E}) \neq \kappa(x)$。

表1. 解释器$\mathbf{L}$的公理。

表1介绍了公理——解释器的形式化性质。其中，成功性、非平凡性和核心性借鉴自 [2], ，其余公理则是原创的。这些公理分为三类：第一类包含任何解释器都可能满足的一般性质，另外两类则定义了不同形式的反事实，下一节将对此进行说明。

成功性 保证每个查询至少有一个解释；非平凡性 则拒绝空解释，因为它们不提供信息；等价性 确保具有相同决策的实例得到相同的解释。该公理适用于生成全局解释的解释器——也就是说，这类解释器关注的是类别，而非输入实例。可行性 指出，解释必须是其决策正在被解释的实例的一部分。我们稍后将看到，对于某些类型的反事实，该公理有助于识别为改变实例的预测而需要更改其值的特征。核心性 更进一步，确保解释仅包含该决策的核心文字。其原理是，缺少一个核心文字必然会导致另一个类别。怀疑有效性 确保解释 E 必然会导致决策发生变化。事实上，所有仅在 E 方面与被解释实例不同的实例，其类别都不同。新颖性 确保解释不与原始实例共享任何文字。其理念是，为了告知用户如何改变输出，解释必须涉及对输入的修改。该公理适用于某些类型的反事实。强有效性 与怀疑有效性的理念相似，因为它确保解释 E 必须导致决策发生变化。事实上，任何包含 E 的实例都会有不同的预测。弱有效性 的要求较低，因为它只要求将原始实例中的新文字替换后得到的实例属于另一个类别。.

如下所示，这些公理并非彼此独立。尽管如此，每条公理在定义不同类型的反事实解释时都起着至关重要的作用。

命题 1. 以下蕴含关系成立。

• 核心性 $\Rightarrow$ 可行性。• 核心性与非平凡性 $\Rightarrow$ 怀疑有效性。• 怀疑有效性 $\Rightarrow$ 非平凡性。• 弱有效性 $\Rightarrow$ 非平凡性。• 强有效性 $\Rightarrow$ 弱有效性。• 新颖性与非平凡性 $\Rightarrow$ 怀疑有效性。

我们提出一个不可能性定理，证明某些公理不相容——任何解释器都无法同时满足所有这些公理。正如我们稍后将展示的，每个相容子集都会产生一种独特的反事实。

定理1. 每个集合$(\mathtt{I}_i)$ 的公理彼此不相容。

• $(\mathtt{I}_1)$成功性、非平凡性和核心性。• $(\mathtt{I}_2)$成功性、可行性和怀疑有效性。• $(\mathtt{I}_3)$成功性、新颖性和强有效性。• $(\mathtt{I}_4)$成功性、非平凡性、可行性和新颖性。• $(\mathtt{I}_5)$成功性、可行性和弱有效性。• $(\mathtt{I}_6)$成功性、非平凡性、等价性和可行性。• $(\mathtt{I}_7)$成功性、非平凡性、等价性和新颖性。

前两项结果未排除的所有公理组合都是相容的；换言之，每种这样的组合至少能由一个解释器满足。

定理 2. 命题 *1* 和定理*1*未排除的所有公理组合，都能由某个解释器满足。

关于这些公理，我们提出了两个根本性问题：

1. 我们能否精确且穷尽地刻画满足给定公理子集的**所有**解释器?
2. 这些公理是否会导出同一类型的反事实?

在接下来的两节中，我们将介绍回答第一个问题的表示定理，并**揭示五种不同类型的反事实**的存在性。这些定理表明，反事实解释分为两种基本形式——必要和充分——每种形式又包含若干不同类型。为清晰起见，我们将分别用一节讨论每种形式。

4 必要理由

第一种反事实可以表述如下：

如果事实并非如此，即 E 不成立，那么决策就会有所不同。

例如，在示例1的实例 x_1 中，如果温度不是炎热的，决策就会不同，这一点可以从表格中验证。这类解释指出了**实例中**应当改变的部分，以避免其当前决策。我们称之为必要理由 ，因为它们是确保当前决策所必需的。本节将研究其全局 和局部版本。前者解释一个类别，而后者针对一个实例。

4.1 全局必要理由

全局必要理由用于解释分类器的行为，即它如何分配类别。对于给定类别，它们指出能够**将该类别与所有其他类别区分开来**的特征–取值组合。在缺少这些组合时，分类器会分配不同的类别。

定义 4。 设$\mathbf{Q} = \langle \mathbb{T}, \kappa, x \rangle$ 为一个查询。 $\kappa(x)$的全局必要理由（*GNR*）是一个集合 $E \in \mathbb{E}(\mathbb{T})$，满足：

$$E \neq \emptyset \quad and \quad \forall y \in \mathbb{F}(\mathbb{T}),\ if\ E \not\subseteq y,\ then\ \kappa(y) \neq \kappa(x).$$

gNec 表示返回所有全局必要理由的解释器。

例 1（续）设$\mathbf{Q}_i$ 为实例 x_i上的查询， $i = 1,2,3$。

- $\mathtt{gNec}(\mathbf{Q}_1) = \{\{(t, \text{hot})\}\}$,
- $\mathtt{gNec}(\mathbf{Q}_2) = \mathtt{gNec}(\mathbf{Q}_3) = \emptyset$.

炎热的温度是海滩选项的特征，因为它使其有别于山和电影院选项。因此，要改变 x_1的决策，必须将 t 的值改为温和或冰冻。相比之下，另外两个查询没有全局必要理由，因为每个特征–值组合都至少出现在被赋予不同类别的两个实例中。例如，修改 x_3 的 a 值并不必然导致类别变化，因为 $\kappa(x_5) = \kappa(x_3)$。

我们给出一个表示定理，用以刻画生成全局必要理由的解释器的**整个族**。它们是同时满足核心性与非平凡性的**仅有的解释器**。

定理 3.解释器$\mathbf{L}$ 满足核心性与非平凡性当且仅当$\mathbf{L} \sqsubseteq \mathtt{gNec}$。

表2总结了 gNec 以及任何满足 $\mathbf{L} \sqsubseteq$ gNec 的解释器$\mathbf{L}$ 所满足或违反的所有性质。

定理4。 *表2*中的性质均成立。

这些解释器违反了成功性，因为定理1表明核心性、成功性和非平凡性彼此不相容。由于 gNec 生成全局解释，它满足等价性：对于被归为同一类别的所有实例，其解释都相同。例如，$\mathtt{gNec}(\mathbf{Q}_1) = \mathtt{gNec}(\mathbf{Q}_6) = \mathtt{gNec}(\mathbf{Q}_7) = \{\{(t,\ \text{炎热})\}\}$，其中 $\mathbf{Q}_i$ 涉及实例 x_i。

总而言之，刻画结果表明，满足核心性和非平凡性公理的解释器族会为某一类别生成全局必要理由。这类反事实解释突出了将给定类别与**所有其他类别**区分开来的因素。

4.2 局部必要理由

现在，我们转而讨论如何解释针对特定实例所作决策的问题。一种自然的方法是考虑与该实例所属类别相关的全局理由。然而，这种方法过于僵化，因为它要求探索整个特征空间，从而大幅降低找到反事实的可能性（例如， x_2， x_3 见例1）。接下来，我们将介绍**怀疑式必要理由**（SNR），这是一种要求较低的替代方法。SNR 是感兴趣实例 x 的子集，用于将该实例与任何会得出不同决策的其他实例 y 区分开来。因此，缺少这些理由便能保证对 x 作出不同的决策。与全局方法不同，局部方法将搜索空间限制在仅在候选原因方面与 x 不同的实例。

定义 5。 令$\mathbf{Q} = \langle \mathbb{T}, \kappa, x \rangle$ 为一个查询。 $\kappa(x)$ 的怀疑式必要理由（*SNR*）是一个集合 $E \in \mathbb{E}(\mathbb{T})$，满足：

$$E \subseteq x \qquad and \qquad \forall y \in x \ominus E,\ \kappa(y) \neq \kappa(x).$$

sNec表示生成所有怀疑式必要理由的解释器。

例 1（续）令 $\mathbf{Q}_i = \langle \mathbb{T},\ \kappa,\ x_i \rangle$。

• $\mathtt{sNec}(\mathbf{Q}_1) = \{\{(t,\ \text{炎热})\},\ x_1\}$。• $\mathtt{sNec}(\mathbf{Q}_2) = \{\{(t,\ \text{温和})\},\ \{(a,\ \text{攀岩})\}\}$。• $\mathtt{sNec}(\mathbf{Q}_3) = \emptyset$。

注意，$x_2 \ominus \{t\} = \{x_1, x_5\}$ $\kappa\ x_2 \neq \kappa\ x_1$ $\kappa\ x_2 \neq (,\ \text{温和}),()$ $\kappa\ y$ $\{t\}$ x_2 $(),()$ $()$. 因此，SNR $(,\ \text{温和})$ 能将其与所有仅在温度上不同于它的实例区分开来**仅**。

注意，全局必要理由都是怀疑式的。因此，如果存在核心文字，那么它们的非空子集都是怀疑式必要理由。但是，解释器可能会返回其他理由，其文字并非实例类别的核心（例如$\mathbf{Q}_2$的情况）。

命题 2。成立的是 $\mathtt{gNec} \sqsubseteq \mathtt{sNec}$。

我们给出一个表示定理，证明这两条公理（可行性、怀疑有效性）与生成怀疑式必要理由的解释器族等价。

定理 5. 解释器 $\mathbf{L}$ 满足可行性和怀疑有效性当且仅当$\mathbf{L} \sqsubseteq \mathtt{sNec}$。

表2总结了 sNec 以及更一般地，任何解释器 $\mathbf{L} \sqsubseteq$ sNec 所满足或违反的公理。请注意，它们满足非平凡性，因此具有**信息性**。然而，它们违反成功性，因此可能无法保证对查询（例如$\mathbf{Q}_3$）给出解释。它们也违反核心性，但满足其较弱版本——怀疑有效性。后者保证无论特征的新取值如何，决策都会发生变化，因此，按照 [15] 中定义的性质，SNRs**有效**。

总而言之，定理5表明，满足可行性和怀疑有效性的解释器集合能够为某个实例生成局部必要理由。这类理由指出了该实例与所有决策不同的其他实例之间的区别因素。

5 充分理由

第二种形式的反事实可以表述如下：

如果 E，决策就会有所不同。

必要理由指出了那些一旦缺失就会改变结果的现有特征，而第二种类型则关注那些一旦存在就会改变结果的缺失特征。例如，“如果 Alice 的年收入是 $\$35K$（而不是实际金额），她就不会被拒绝贷款。”我们引入充分理由这一概念——即直接导致不同决策的特征值——而不是指出实例中哪些部分必须修改。

5.1 全局充分理由

我们首先分析分类器的全局行为，从而分析其各类别的全局行为。如前所述，全局必要原因能够揭示一个类别与其他类别的区别；然而，这类原因并非总是存在，示例1中的实例 x_2 和 x_3 便说明了这一点。为此，我们引入全局充分原因的概念，用于避免某个类别——这些因素的存在会促使分类器将实例归入其他类别。

	必要理由				充分理由					
公理	$\mathbf{L} \sqsubset$ gNec	gNec $\mathbf{L} \sqsubset$ sNec	sNec $\mathbf{L} \sqsubset$	gSuf	gSuf $\mathbf{L} \sqsubset$	sSuf	sSuf $\mathbf{L} \sqsubset$	cSuf	cSuf	
成功性	×	×	×	×	–	✓	×	×	–	✓
非平凡性	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
等价性	–	✓	–	×	–	✓	–	×	–	×
可行性	✓	✓	✓	✓	–	×	–	×	–	×
核心性	✓	✓	–	×	–	×	×	×	×	×
怀疑有效性	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
新颖性	–	×	–	×	–	×	✓	✓	✓	✓
强有效性	–	×	–	×	✓	✓	✓	✓	–	×
弱有效性	–	×	–	×	✓	✓	✓	✓	✓	✓

表2. 解释器的形式比较。“–”表示该公理仅由该类的部分实例满足，而非 all.

定义 6. 设$\mathbf{Q} = \langle \mathbb{T}, \kappa, x \rangle$ 是一个查询。针对 $\kappa(x)$ 的全局充分原因（*GSR*）是一个集合 $E \in \mathbb{E}(\mathbb{T})$，满足：

$$\forall y \in \mathbb{F}(\mathbb{T}) \textit{ with } E \subseteq y,\ \kappa(y) \neq \kappa(x).$$

令gSuf 为生成所有全局充分理由的解释器。

例 1（续） 令 $\mathbf{Q}_i = \langle \mathbb{T}, \kappa, x_i \rangle, i = 1,2,3$。

• gSuf($\mathbf{Q}_1$ $= \{\{t$ $\} \{t$ $\}\} \cup \{$(,温和),(,冰冻)$\}$。
• gSuf($\mathbf{Q}_2$ $= \{\{t$ $\} \{a$ $\}\} \cup \{$(,炎热),(,阅读)$\}$。 • gSuf($\mathbf{Q}_3$) $= \{\{$(t,炎热)$\}\} \cup \{x_1, x_2, x_4, x_6, x_7\}$。

注意，$\mathbf{Q}_1$ 有 8 个全局充分理由；第一个是：如果温度是温和的，决策就会与海滩不同。

下述表示定理表明，只有满足强有效性的解释器才能生成全局充分理由。

定理 6。 解释器**L**满足强有效性当且仅当$\mathbf{L} \sqsubseteq$ gSuf。

表2总结了 gSuf 以及任何满足$\mathbf{L} \sqsubseteq$ gSuf 的解释器在所有公理方面的表现。它证实了两种形式的反事实——必要反事实和充分反事实——传达了不同类型的信息，因为它们满足不同的公理子集。与 gNec 不同，gSuf 保证对任何查询至少有一个非空解释，但违反可行性。

总而言之，定理 6 表明，强有效性公理催生了一种新型全局反事实——这类反事实强调特征—取值组合，其存在能够保证避免属于某一给定类别。

5.2 局部充分理由

现在让我们解释个体实例的决策。我们区分两种类型的解释：怀疑型和确信型。

5.2.1 怀疑充分理由*(SSR)*

它们突出了被归入与所考察实例不同类别的实例所具有的特征。因此，它们充当其他类别的区分性特征，并且在所关注的实例中必须不存在。 consequently，怀疑充分理由完全由实例中尚未存在的**新特征值**组成。

定义 7。 令$\mathbf{Q} = \langle \mathbb{T}、\ \kappa、\ x \rangle$ 为一个查询。针对 κ（x） 的怀疑充分理由 （*SSR*）是一个集合 $E \in \mathbb{E}$（$\mathbb{T}$），满足：

$$E \cap x = \emptyset \quad \textit{and} \quad \forall y \in \mathbb{F}(\mathbb{T}) \textit{ s.t. } E \subseteq y,\ \kappa(y) \neq \kappa(x).$$

sSuf表示生成所有怀疑充分理由的解释器。

Ex. 1 (Cont) Let $\mathbf{Q}_i = \langle \mathbb{T}, \kappa, x_i \rangle, i = 1, 2, 3.$

• sSuf($\mathbf{Q}_1$) $= \{\{$(t, 温和)$\}、\ \{$(t, 冰冻)$\}\} \cup \{x_3, x_4, x_8, x_9\}$。 •
sSuf($\mathbf{Q}_2$) $= \{\{$(t, 炎热)$\}、\ \{$(a, 阅读)$\}\} \cup \{x_3, x_6, x_7\}$。 •
sSuf($\mathbf{Q}_3$) $= \{\{$(t, 炎热)$\}\} \cup \{x_1, x_2, x_7\}$。

例如，考虑理由 {(t, 温和)}。这一特征—取值对刻画了被标记为山或电影院的实例。没有任何被标记为海滩的实例具有温和的温度。

下一个表示定理确立了两条公理——新颖性和强有效性——与生成 SSR 的解释器集合之间的一一对应关系。

定理7. 解释器 **L**满足新颖性和强有效性当且仅当 $\mathbf{L} \sqsubseteq$ sSuf。

表2显示，这第四类反事实与局部必要理由不同。不过，这两类都可能不存在。

总而言之，上述刻画表明，强有效性和新颖性定义了第四类反事实：它们基于缺失的特征–取值组合，且总会导致不同的决策。

5.2.2 确信型充分理由 （*CSR*）。

作为限制最少的一类理由，宽容充分理由是一些特征–取值组合，它们能够区分某个实例与另一个被划分为不同类别的**实例**。与怀疑充分理由不同，存在宽容充分理由并不一定能保证决策发生变化。在本文的示例中，滑雪是实例 x_3的一个宽容充分理由，因为将 a 的取值改为滑雪会得到 x_4，后者获得了不同的决策。不过，滑雪不是怀疑充分理由，因为它也出现在 x_9中，而 x_9的类别与 x_3相同。

定义 8. 设$\mathbf{Q} = \langle \mathbb{T}, \kappa, x \rangle$ 为一个查询。针对 $\kappa(x)$ 的宽容充分反理由是满足以下条件的 $y \setminus x$： $y \in \mathbb{F}(\mathbb{T})$ 且 $\kappa(y) \neq \kappa(x)$。cSuf 是生成所有宽容充分理由的解释器。

例 1（续） 设 $\mathbf{Q}_i = \langle \mathbb{T}, \kappa, x_i \rangle, i = 1,2,3$。

• cSuf($\mathbf{Q}_1$) =sSuf($\mathbf{Q}_1$).
• cSuf($\mathbf{Q}_2$) $= \{\{$(t, 炎热)$\}, \{$(t, 冰冻)$\}, \{$(a, 阅读)$\},$
$\{a$ $\}\} \cup \{y \in \mathbb{F}\ \mathbb{T}\ \mid y \cap x_2 = \emptyset$ $\kappa\ y\ \neq$ (,滑雪) ( )和()山}。

• cSuf(Q 3 = {{ t } { a }}∪{y ∈ F T | y∩x3 =) (, 炎热),(, 滑雪) ( ) ∅ κ y ≠ } 和( )电影院。

注意，{(a，滑雪)} 不是 $\mathbf{Q}_3$ 的 SSR，因为滑雪并不代表山或海滩。事实上，κ 会为实例 $x_9 = ((t,$ 温和)，$(a,$ 滑雪)) 预测电影院（x_3的类别）。

接下来，我们将介绍一个表示定理，证明只有返回宽容充分理由的解释器才满足新颖性和弱有效性这两条公理。

定理 8. 解释器 $\mathbf{L}$ 满足新颖性和弱有效性当且仅当 $\mathbf{L} \sqsubseteq$ cSuf。

现在，我们引入**忠实排序**这一概念——它是定义在部分赋值集合上的预序，会严格优先考虑那些为实例引入新特征值的赋值，前提是这些修改会导致决策发生变化。

定义 9. 忠实排序是一种函数$\mathbf{S}$，它将每个查询$\mathbf{Q} = \langle \mathbb{T}, \kappa, x \rangle$ 映射为定义在 $\mathbb{E}(\mathbb{T})$ 上的预序 $\succeq_{\mathbf{Q}}$，使得 $\forall E, E' \in \mathbb{E}(\mathbb{T})$，如果：

- $E \cap x = \emptyset$ 和 $\kappa\,(x_{\downarrow E}) \neq \kappa\,(x)$ 以及
- $E' \cap x \neq \emptyset$ 或 $\kappa\,(x_{\downarrow E'}) = \kappa\,(x)$，

那么 $E \succ_{\mathbf{Q}} E'$。

接下来，我们将介绍第二个表示定理，它将成功性、新颖性和弱有效性公理与宽容充分理由联系起来。该定理确立了这些公理与一种基于部分赋值预序的解释策略之间的等价性。具体而言，满足这些公理的解释器会对所有部分赋值进行排序，从最低偏好到最高偏好，并选取排名最高的部分赋值作为解释。

定理 9. 解释器$\mathbf{L}$ 满足成功性、新颖性和弱有效性当且仅当存在一个忠实排序$\mathbf{S}$，它将每个查询$\mathbf{Q} = \langle \mathbb{T}, \kappa, x \rangle$ 映射到（$\mathbb{T}$）上的一个预序 $\succeq_{\mathbf{Q}}$，使得：

$$\mathbf{L}(\mathbf{Q}) = \max(\mathbb{E}(\mathbb{T}), \succeq_{\mathbf{Q}}).$$

上述结果刻画了定理 8 中所确定的解释器子族$\mathbf{L}$ ⊑cSuf，该子族保证至少存在一个 CSR。在后续章节中，我们将展示，这种以忠实排序表述的刻画有助于比较生成宽容充分理由的现有解释器。

下一个结果将三类充分理由（GSR、SSR、CSR）联系起来，并表明每个局部解释都是某个全局解释的子集。

命题 3. 设 $\mathbf{Q} = \langle \mathbb{T},\ \kappa,\ x \rangle$ 为一个查询。

• sSuf ⊑gSuf。•
sSuf ⊑cSuf。
• 对于任意 $E \in \mathtt{cSuf}(\mathbf{Q})$，$\exists E' \in \mathtt{gSuf}(\mathbf{Q})$ 使得 $E = E' \setminus x$。
• 对于任意 $E \in \mathtt{gSuf}(\mathbf{Q})$，$E \setminus x \in \mathtt{cSuf}(\mathbf{Q})$。

表2中总结的性质将这第五类反事实与其他四类区分开来。与 SSR 不同，信任式充分理由必然存在。因此，定义明确的解释器能够为每个查询提供一个结果。

6 以人为中心的类型评估

认知科学研究表明，人们通常更喜欢简短精炼的解释 [7, 48, 8]。我们根据以下比较标准，考察五类反事实与这些需求的契合程度。

存在性：反事实是否保证存在。**紧凑性：**以简洁方式传达多条信息的能力。 **大小：**一个反事实中的文字数量。**数量：**针对一个查询所提供的解释数量。**区分能力：**反事实在实例或类别之间进行区分的能力。

表2显示，必要理由和怀疑型充分理由可能并非总是存在，而（全局、确信型）充分理由则保证对任何查询至少存在一个解释。虽然理由可能不存在这一点可以被视为缺点，[15] 强调，确信型充分理由中的特征可能并非可操作 （例如，年龄无法改变）。因此，生成由待修改特征组成的宽容充分理由的解释器 $\mathbf{L} \sqsubseteq$ cSuf可能会遇到反事实不存在的情况。

当必要理由（无论是局部还是全局）存在时，它们表示这样一些特征-取值组合：仅这些组合的缺失就会导致类别发生变化，而不论为这些特征赋予了哪些新取值。因此，无须列出所有可能导致决策改变的变化。这使必要理由能够以**简洁**的方式传达多项信息——变化的全部范围，因此对人们尤其有吸引力。相比之下，充分理由并不简洁，因为它们会明确列举所有可能的变化。

由于其紧凑性，必要理由通常**数量更少**，不及充分理由。在贯穿示例中，gNec 对查询$\mathbf{Q}_1$, 只返回一个解释，而 gSuf 则产生八个全局充分理由。事实上，出现在全局必要理由中的某一特征的每个备选值都会产生不同的全局充分理由。全局必要理由也比全局充分理由**更短**，因为每个核心文字都提供一个最小解释。

全局反事实是**有区分性的**，因为它们捕捉那些将一个类别与所有其他类别区分开来的特征–值对。同样，怀疑式（必要、充分）理由识别出能使一个实例区别于其他实例的组合。然而，宽容充分理由的区分性较低，因为它们可能包含同样出现在具有相同结果的实例中的配对。例如，(t,冰冻)是 x_2 的一个宽容充分理由，但也属于 x_4，后者与 x_2具有相同的决策。
总之，必要理由只要存在，似乎就比充分理由更符合人类偏好。

7 相关工作

大量研究探讨了反事实解释（参见[15] 中的最新综述）。一些研究提出了新的解释器，而另一些则着重于确定反事实及其生成解释的解释器应满足的理想属性。下面我们将回顾这两方面的工作。

7.1 反事实类型与解释器

尽管关于反事实解释的研究成果丰富，但绝大多数都聚焦于**局部** 反事实（例如，[13, 14, 25, 32, 36, 39, 42, 46] x ）。此外，给定一个输入，这些

研究会生成能够获得不同决策的"最近"实例 y。各类方案的主要区别在于如何定义接近度。有些方案强调**稀疏性**，力求尽量减少需要更改的特征数量；另一些则优先考虑**邻近性**，只要依据指定的**距离度量**（如曼哈顿距离（L_1 范数）或*Gower* 距离）计算出的总体变化较小，便允许进行更广泛的更改（更多示例见 [46]）。我们证明，所有这些解释器都会生成**宽容充分理由**，因此都是 CSR 族中的特定实例。换言之，每个这样的解释器，例如$\mathbf{L}$，都满足以下条件：$\mathbf{L} \sqsubseteq$ cSuf。此外，我们还刻画了每个解释器所定义的忠实排序。

有两种促进稀疏性的解释器。第一种生成对比解释 [11]—**为了获得不同的决策，必须改变的最小特征集合**。为便于比较，我们介绍与之对应的解释器，其解释是文字集合（而非特征集合）。

定义 10。我们将$\mathbf{L_{wf}}$ 定义为这样一种解释器：对于任意查询$\mathbf{Q} = \langle \mathbb{T}, \kappa, x \rangle$ 和 $E \in \mathbb{E}(\mathbb{T})$，$E \in \mathbf{L_{wf}}(\mathbf{Q})$ 当且仅当：

- $\kappa(x_{\downarrow E}) \neq \kappa(x)$,
- $\nexists E' \in \mathbb{E}(\mathbb{T})$ 使得 $\kappa(x_{\downarrow E'}) \neq \kappa(x)$, 特征$(E') \subset$ 特征(E)（其中 特征(E) 是 E 中涵盖的特征集合）。

例 1（续）令 $\mathbf{Q}_i = \langle \mathbb{T}, \kappa, x_i \rangle, i = 1,2,3$。

- $\mathbf{L_{wf}}(\mathbf{Q}_1) = \{\{(t,\ \text{温和})\},\ \{(t,\ \text{冰冻})\}\}$,
- $\mathbf{L_{wf}}(\mathbf{Q}_2) = \{\{(t,\ \text{炎热})\},\ \{(a,\ \text{阅读})\}\}$,
- $\mathbf{L_{wf}}(\mathbf{Q}_3) = \{\{(t,\ \text{炎热})\}\}$。

下面我们定义与$\mathbf{L_{wf}}$相对应的预序。

定义 11。设$\mathbf{Q} = \langle \mathbb{T}, \kappa, x \rangle$ 是一个查询，δ 是$\mathbb{E}(\mathbb{T})$ 上的权重函数，使得对于任意 $E \in \mathbb{E}(\mathbb{T})$：

$$\delta(E) = \begin{cases} 1 & \text{if } E \cap x = \emptyset \text{ and } \kappa(x_{\downarrow E}) \neq \kappa(x) \\ +\infty & \text{otherwise} \end{cases}$$

$\succeq^f_{\mathbf{Q}}$ 是 $\mathbb{E}(\mathbb{T})$ 上的一个预序，使得 $\forall E, E' \in \mathbb{E}(\mathbb{T})$，$E \succeq^f_{\mathbf{Q}} E'$ 当且仅当$(\delta(E) = \delta(E')$ 且特征$(E) \subseteq$ 特征$(E'))$ 或者 $\delta(E) < \delta(E')$

我们证明$\mathbf{L_{wf}}$ 是CSR族的一个实例。

定理10。以下性质成立：

- 将每个查询 $\mathbf{S}$ $\mathbf{Q}$ 映射到 $\succeq^f_{\mathbf{Q}}$（在 $\mathbb{E}(\mathbb{T})$ 上）的函数是一个忠实排序。
- 对于任意查询 $\mathbf{Q}$，$\mathbf{L_{wf}}(\mathbf{Q}) = \max(\mathbb{E}(\mathbb{T}) \succeq^f_{\mathbf{Q}})$。

第二个解释器在 [15], 中提出，它通过选择**基数最小**的解释来体现稀疏性。

定义 12。我们将$\mathbf{L}_c$ 定义为这样的解释器：对于任意查询$\mathbf{Q} = \langle \mathbb{T}, \kappa, x \rangle$ 和 $E \in \mathbb{E}(\mathbb{T})$，当且仅当满足以下条件时，$E \in \mathbf{L}_c(\mathbf{Q})$：

- $\kappa(x_{\downarrow E}) \neq \kappa(x)$,
- $\nexists E' \in \mathbb{E}(\mathbb{T})$ *s.t* $\kappa(x_{\downarrow E'}) \neq \kappa(x)$ *and* $|E'| < |E|$.

备注： 不难证明$\mathbf{L}_c \sqsubseteq \mathbf{L_{wf}}$。

例 1（续） 对于每个 $i = 1,\ 2,\ 3$，$\mathbf{L}_c(\mathbf{Q}_i) = \mathbf{L_{wf}}(\mathbf{Q}_i)$。

该解释器根据宽容充分理由的基数对其进行排序。

定义13. 令$\mathbf{Q} = \langle \mathbb{T}, \kappa, x \rangle$ 为一个查询，σ 为$\mathbb{E}(\mathbb{T})$ 上的权重函数，使得对于任意 $E \in \mathbb{E}(\mathbb{T})$：

$$\sigma(E) = \begin{cases} |E| & \text{if } E \cap x = \emptyset \text{ and } \kappa(x_{\downarrow E}) \neq \kappa(x) \\ +\infty & \text{otherwise} \end{cases}$$

我们将 $\succeq^c_{\mathbf{Q}}$ 定义为 $\mathbb{E}(\mathbb{T})$上的一个预序，使得：

$$\forall E, E' \in \mathbb{E}(\mathbb{T}),\ E \succeq^c_{\mathbf{Q}} E' \text{ iff } \sigma(E) \leq \sigma(E').$$

我们证明$\mathbf{L}_c$ 是 CSR 族的一个实例 .

定理 *11. </style}*以下性质成立：

- 将每个查询$\mathbf{Q}$ 映射到$\succeq^c_{\mathbf{Q}}$ 上的 $\mathbb{E}(\mathbb{T})$ 的函数$\mathbf{S}$ 是一个忠实排序。
- 对于任意查询$\mathbf{Q}$，$\mathbf{L}_c(\mathbf{Q}) = \max(\mathbb{E}(\mathbb{T}) \succeq^c_{\mathbf{Q}})$。

现有方法使用距离度量比较候选反事实（包括[3, 12, 14, 25, 32, 36, 39, 42, 46] 中提出的方法），**都**按以下方式定义。

定义 14. </style}设$\mathbf{Q} = \langle \mathbb{T}, \kappa, x \rangle$ 是一个查询，$E \in \mathbb{E}(\mathbb{T})$，且$\mathsf{d}$ 是 $\mathbb{F}(\mathbb{T})$上的距离度量。基于d 的解释器是一个函数$\mathbf{L}_{\mathsf{d}}$，满足 $E \in \mathbf{L}_{\mathsf{d}}(\mathbf{Q})$ 当且仅当：

- $\kappa(x_{\downarrow E}) \neq \kappa(x)$,
- $\nexists E' \in \mathbb{E}(\mathbb{T}), \kappa(x_{\downarrow E'}) \neq \kappa(x) \wedge \mathsf{d}(x_{\downarrow E'}, x) < \mathsf{d}(x_{\downarrow E}, x)$

现有的基于距离的解释器在距离度量 d 的选择上各不相同，因此可能产生与$\mathbf{L_{wf}}$ 和$\mathbf{L}_c$所产生的解释不同的结果，如下所示。

例 1（续）令 $y = \{(t,\ \text{冰冻}), (a,\ \text{攀岩})\}$。假设距离度量 d 满足 $\mathsf{d}(\{\{t\}\}, \{\}) < \mathsf{d}(\{\}, \{\})$。那么，$\mathbf{L}(\mathbf{Q})(\{\}, \text{温和})$。

我们定义这些解释器所产生的预序。

定义 15。令$\mathbf{Q} = \langle \mathbb{T}, \kappa, x \rangle$ 为一个查询，d 为 $\mathbb{F}(\mathbb{T})$ 上的距离度量，γ 为 $\mathbb{E}(\mathbb{T})$ 上的权重函数，满足 $\forall E \in \mathbb{E}(\mathbb{T})$，

$$\gamma(E) = \begin{cases} \mathsf{d}(x_{\downarrow E}, x) & \text{if } E \cap x = \emptyset \wedge \kappa(x_{\downarrow E}) \neq \kappa(x) \\ +\infty & \text{otherwise} \end{cases}$$

我们将$\succeq_{\mathbf{Q},\mathsf{d}}$ 定义为 $\mathbb{E}(\mathbb{T})$上的一个预序，使得：

$$\forall E, E' \in \mathbb{E}(\mathbb{T}),\ E \succeq_{\mathbf{Q},\mathsf{d}} E' \text{ iff } \gamma(E) \leq \gamma(E').$$

我们证明$\mathbf{L}_{\mathsf{d}}$ 是 CSR 族的一个实例。

定理12。令d 为距离度量。

- 将每个查询$\mathbf{Q}$ 映射到$\succeq_{\mathbf{Q},\mathsf{d}}$ 上的函数$\mathbf{S}$ $\mathbb{E}(\mathbb{T})$ 是忠实排序。
- 对于任意查询$\mathbf{Q}$，$\mathbf{L}_{\mathsf{d}}(\mathbf{Q}) = \max(\mathbb{E}(\mathbb{T}) \succeq_{\mathbf{Q},\mathsf{d}})$。

在刘易斯式反事实中（最初在开创性论文[29],then in [50, 1]中提出），变量集合的变化必须很小。这可以用以下函数$\mathbf{L}_{\mathsf{d},\tau}$来建模，该函数也是宽容充分理由（即$\mathbf{L}_{\mathsf{d},\tau} \sqsubseteq$ cSuf）这一类的一个实例。

定义 *16*。令 $Q = \langle \mathbb{T}, \kappa, x \rangle$ 为一个查询，$E \in \mathbb{E}(\mathbb{T})$，且 d 是 $\mathbb{F}(\mathbb{T})$ 上的距离度量。基于 $(d\ \tau)$ 的阈值解释器是一个函数 $L_{\mathsf{d},\tau}$，满足 $E \in L_{\mathsf{d},\tau}(Q)$ 当且仅当 $\kappa(x_{\downarrow E}) \neq \kappa(x)$ 且 $d(x_{\downarrow E}, x) < \tau$。

有三项值得注意的研究探讨了**必要理由**。其中两项由 [20, 30], 提出了函数，用于为实例生成子集极小的必要理由，但违反了

怀疑式有效性公理。如果存在核心文字，或理论中的所有特征都是二元的，它们给出的理由就是怀疑式必要理由。 然而，在一般情况下，这些理由只能保证存在某种变化，却没有说明具体是哪一种变化。在这种情况下，它们对人类用户毫无用处。第三项研究由 [2], 开展，研究了全局反事实。该研究提出了一些公理，包括核心性、成功性和非平凡性，并定义了两个解释器，分别记作$\mathbf{L}_1$ 和$\mathbf{L}_2$: 一个返回单一理由，即分配给该实例的类别的全部核心文字；另一个则返回非空、子集极小的必要理由。这两个解释器都会得出全局必要理由，即$\mathbf{L}_1 \sqsubseteq$ gNec和$\mathbf{L}_2 \sqsubseteq$ gNec。我们的工作更具一般性，探讨了范围更广的反事实类型，包括全局必要理由、怀疑式必要理由、GSR、SSR和CSR，并取得了更深入的理论成果，尤其是基于六条全新公理提出了一系列表示定理。

必要理由这一概念在 [13] 中被引入，定义为所有能够保证实例类别的部分赋值的交集。该理由是唯一的，可能为空，移除它也不一定会改变决策。因此，相应的解释器无法提供足够的信息，帮助用户了解如何改变预测，因而不符合反事实解释的定义。

7.2 反事实（解释器）的性质

已有多篇论文研究了解释的性质 [5,17, 44, 49, 47] ，而 [31, 23, 39, 40, 43, 26, 45]提出了反事实的各种性质。这些论文聚焦于宽容充分理由（CSRs），并讨论了各种定量性质，这些性质用于定义函数，从 cSuf 返回的反事实中选出一个子集。性质示例包括最小性、相似性、合理性、区分能力、可操作性、因果性以及公平性。最小性保证基数最小；相似性要求待解释实例与反事实之间的距离最小；合理性则要求 CSR 中的新取值应与训练阶段所见实例的取值相似 [31, 23, 39, 40, 43]。区分能力是一种主观性质，指反事实应有助于弄清楚为何能够得到不同的结果。可操作性保证特征是可操作的，即可以被更改。这一性质非常重要，但可能导致成功性的违反。因果性保证反事实保留属性之间已知的因果关系 [43]。如果反事实不涉及受保护特征，则称其公平 [26, 45]。解释的公平性与决策的公平性密切相关；事实上，前者可用于评估后者 [19]。

集合反事实解释的性质也已被提出，即多样性和稳定性，也称为鲁棒性 [22]。多样性要求所选择的反事实应当彼此不同，例如，涉及不同的属性 [28, 36, 39, 41]。稳定性 [16] 则认为，相似的实例应获得相似的解释。

备注： 第4节和第5节介绍了五大类解释器，每类解释器都会生成特定类型的反事实（GNR、SNR、GSR、SSR、CSR）。上述性质**补充**了这一分类，有助于在每一类中识别“合理”的解释器。例如，可以定义一种解释器，用于生成满足子集最小、可操作、多样且公平的SNR。

最后，虽然 [4]研究了**“为什么”**问题的解释器**固定**类别的性质，并分析了三种分类器的决策复杂度，但我们的工作聚焦于**“为什么不”**问题。我们提出了新的公理，用以刻画五类反事实解释器，并分析任意分类器下的解释复杂度。

8 计算复杂度

本节分析与反事实解释器$\mathbf{L}$相关的两个基本计算问题:

- DECIDEEXP($\mathbf{L}$) 是这样一个判定问题：其输入为查询$\mathbf{Q}$ 和解释 E, 用于确定 $E \in \mathbf{L}(\mathbf{Q})$ 是否成立。
- FINDEXP($\mathbf{L}$)以查询$\mathbf{Q}$ 作为输入，并返回一个元素 $E \in \mathbf{L}(\mathbf{Q})$（若不存在则返回“none”）。

这些问题的复杂度取决于可能的分类器族 κ。具体而言，我们假设 κ是一个已知的布尔公式。如果复杂度取决于域大小，我们也会考虑更大的定义域。

定理 13。 考虑查询$\mathbf{Q} = \langle \mathbb{T}, \kappa, x \rangle$，其中 $\mathbb{T}$ 由 n 个布尔特征组成，且 κ 是一个布尔公式。

1. 当$\mathbf{L}$ 为cSuf、sNec或$\mathbf{L}_{\mathbf{wf}}$时，问题DECIDEEXP($\mathbf{L}$) 属于 $O(n)$ 。
2. 当$\mathbf{L}$ 为gNec、gSuf、sSuf、$\mathbf{L}_{\mathbf{c}}$或$\mathbf{L}_d$时，问题DECIDEEXP($\mathbf{L}$)是 *co-NP* 完全。

对于域大小任意的情形，当$\mathbf{L}$ 为cSuf 或$\mathbf{L}_{\mathbf{wf}}$时，问题DECIDEEXP($\mathbf{L}$) 属于 $O(n)$ 。对于非布尔域上的查询，当$\mathbf{L}$ 为sNec时，问题 DECIDEEXP($\mathbf{L}$) 是 *co-NP* 完全。

我们研究生成解释的复杂度。

定理 14. 考虑查询$\mathbf{Q} = \langle \mathbb{T}, \kappa, x \rangle$，其中 $\mathbb{T}$ 由 n 布尔特征组成，且 κ 是一个布尔公式。

1. 问题FINDEXP($\mathbf{L}$) 在$\mathbf{L}$ 为sSuf时属于 $O(n)$ 。*2.* 问题FINDEXP($\mathbf{L}$) 在 $\mathbf{L}$ 为gNec、cSuf、sNec、gSuf、$\mathbf{L}_{\mathbf{c}}$、$\mathbf{L}_d$或$\mathbf{L}_{\mathbf{wf}}$时是 *NP* 难。在这些情况下，除$\mathbf{L}_{\mathbf{c}}$ 和$\mathbf{L}_d$外，最多调用 n 次 *SAT* 预言机即可求解。

对于非布尔域上的查询，FINDEXP($\mathbf{L}$) 在sSuf时是 *NP* 难的。

定理 13 和 14 为可表示为布尔公式的白盒分类器给出了复杂度上界。不过需要指出，对于某些特定类型的分类器，例如决策树或单调函数，某些解释查询可以在多项式时间内完成 [4, 11, 18, 21, 33]。

9 结论

本文提出了首个针对反事实解释的公理化框架。我们定义了一组公理，并建立了刻画五种反事实类型的表示定理，其中三种是新提出的：SNR、GSR 和 SSR。这些类型可归为两种基本形式——必要理由和充分理由。必要理由简洁且信息丰富，但不一定总是存在；充分理由则保证存在，但数量可能很多，尤其是在特征域较大的情况下。除布尔特征空间等受限情形外，这两种形式通常都难以计算。

这项工作为未来研究开辟了若干方向。一个方向是针对每种反事实，刻画能够生成该类解释且满足可操作性、最小性、公平性和多样性等既定性质的解释器。另一个方向是研究新引入的反事实类型的鲁棒性。

致谢

我们的工作得到人工智能跨学科研究所 ANITI 的支持。该研究所由法国2030计划资助，资助协议编号为 ANR-23-IACL-0002；此外，我们还获得了法国国家研究署项目 ForML（ANR-23-CE25-0009）的支持。

参考文献

[1] C. Aguilera-Ventura、A. Herzig、X. Liu 和 E. Lorini。基于基础距离的反事实推理。载于《第20届知识表示与推理原则国际会议论文集》，知识表示与推理会议，第 2–11 页，2023年。[2] L. Amgoud 和 J. Ben-Naim。可解释性的公理化基础。载于 IJCAI，第 636–642 页，2022。[3] G. Audemard、J. Lagniez 和 P. Marquis。提升回归树对比性解释的计算。载于第27届欧洲人工智能会议，欧洲人工智能会议，第 392 卷，第 1083–1091 页。IOS Press，2024。[4] S. Bassan、G. Amir 和 G. Katz。局部与全局可解释性：计算复杂度视角。载于国际机器学习大会，《机器学习研究论文集》第 235 卷，第 3133–3167 页，2024。网址：https://openreview.net/forum?id=veEjiN2w9F。[5] F. Bodria、F. Giannotti、R. Guidotti、F. Naretto、D. Pedreschi 和 S. Rinzivillo。黑箱模型解释方法的基准测试与综述。《数据挖掘与知识发现》，37(5)：1719–1778，2023年。[6] R. Byrne。反事实思维。《心理学年度评论》，67：135–157，2016。[7] R. Byrne。可解释人工智能（可解释人工智能）中的反事实：人类推理的证据。载于《第二十八届国际人工智能联合会议论文集》，IJCAI，第 6276–6282 页，2019年。[8] J. Cassens、L. Habenicht、J. Blohm、R. Wegener、J. Korman、S. Khemlani、G. Gronchi、R. M. J. Byrne、G. Warren、M. S. Quinn 和 M. T. Keane。人类思维中的解释。载于《第43届认知科学学会年会论文集》。cognitivesciencesociety.org，2021年。[9] J. Chen、Y. Wang、J. Wang、X. Xie、J. Hu、Q. Wang 和 F. Xu。通过反事实推理理解多智能体系统中各个智能体的重要性。载于人工智能促进协会年会-25，由人工智能促进协会主办，第 15785–15794 页。AAAI Press，2025。[10] M. C. Cooper 和 L. Amgoud。约束下分类器的溯因解释：复杂度与性质。载于欧洲人工智能会议，第 372 卷，第 469–476 页。IOS Press，2023年。[11] M. C. Cooper 和 J. Marques-Silva。解释分类器决策的可处理性。《人工智能》，316：103841，2023年。doi：https://doi.org/10.1016/j.artint.2022.103841。[12] S. Dandl、C. Molnar、M. Binder 和 B. Bischl。多目标反事实解释。载于《自然界并行问题求解——PPSN XVI》，第 448–469 页。Springer International Publishing，2020年。[13] A. Darwiche 和 A. Hirth。决策背后的原因。载于第24届欧洲人工智能会议，欧洲人工智能会议，第 325 卷，第 712–720 页。IOS Press，2020年。[14] A. Dhurandhar、P. Chen、R. Luss、C. Tu、P. Ting、K. Shanmugam 和 P. Das。基于缺失信息的解释：迈向包含相关负例的对比性解释。载于神经信息处理系统年会，神经信息处理系统，第 590–601 页，2018年。[15] R. Guidotti。反事实解释及其寻找方法：文献综述与基准测试。《数据挖掘与知识发现》，38(5)：2770–2824，2024。doi：10.1007/S10618-022-00831-6。[16] R. Guidotti 和 S. Ruggieri。可解释模型的稳定性。载于《2019年国际神经网络联合会议》，第 1–8 页，2019年。[17] R. Guidotti、A. Monreale、S. Matwin 和 D. Pedreschi。通过在潜在特征空间中学习图像范例来解释黑箱。载于《数据库中的机器学习与知识发现：欧洲会议》，ECML PKDD，第 189–205 页，2019年。[18] X. Huang、Y. Izza、A. Ignatiev、M. C. Cooper、N. Asher 和 J. Marques-Silva。d-DNNF分类器的可处理解释。载于人工智能促进协会年会，第 5719–5728 页。AAAI Press，2022。doi：10.1609/AAAI.V36I5.20514。网址：https://doi.org/10.1609/aaai.v36i5.20514。[19] A. Ignatiev、M. C. Cooper、M. Siala、E. Hebrard 和 J. Marques-Silva。迈向机器学习中的形式化公平性。载于 CP，《计算机科学讲义》第 12333 卷，第 846–867 页，2020。[20] A. Ignatiev、N. Narodytska、N. Asher 和 J. Marques-Silva。从对比解释到溯因解释，再回到对比解释。载于 M. Baldoni 和 S. Bandini 编，AIxIA 2020，计算机科学讲义第 12414 卷，第 335–355 页。Springer，2020年。[21] Y. Izza、A. Ignatiev 和 J. Marques-Silva。论如何解决决策树中的解释冗余问题。J. Artif. Intell. Res.，75:261–321，2022。[22] J. Jiang、F. Leofante、A. Rago 和 F. Toni。机器学习中的鲁棒反事实解释：一项综述。载于第三十三届国际人工智能联合会议论文集，IJCAI 2024，韩国济州，2024年8月3日至9日，第 8086–8094 页。ijcai.org，2024。[23] K. Kanamori、T. Takagi、K. Kobayashi 和 Y. Ike。反事实解释树：借助决策树实现透明且一致的可行动补救。载于人工智能与统计国际会议，AISTATS，第 151 卷，第 1846–1870 页。机器学习研究会论文集，2022。[24] A.-H. Karimi、G. Barthe、B. Schölkopf 和 I. Valera。算法补救综述：对比解释与后果性建议。ACM Computing Surveys，55(5)，2022。[25] M. T. Keane、E. M. Kenny、E. Delaney 和 B. Smyth。要是我们有更好的反事实解释就好了：反事实可解释人工智能技术评估中亟待弥补的五项关键缺陷。载于第三十届国际人工智能联合会议论文集，IJCAI，第 4466–4474 页，2021。[26] M. J. Kusner、J. R. Loftus、C. Russell 和 R. Silva。反事实公平性。载于神经信息处理系统大会，第 4066–4076 页，2017。[27] F. Leofante 和 A. Lomuscio。迈向适用于人类—神经多智能体系统的鲁棒对比解释。载于2023年国际自主智能体与多智能体系统会议论文集，AAMAS，第 2343–2345 页。美国计算机协会，2023。[28] F. Leofante 和 N. Potyka。通过多样性提升反事实鲁棒性。载于人工智能促进协会年会，第 38(19) 卷，第 21322–21330 页，204年。[29] D. Lewis。《反事实》。哈佛大学出版社，马萨诸塞州剑桥，1973年。[30] X. Liu 和 E. Lorini。分类器系统中解释的统一逻辑框架。逻辑与计算杂志，33 515，2023 2 485( ): – 。[31] A. V. Looveren 和 J. Klaise。由原型引导的可解释反事实解释。载于 ECML PKDD，第 12976 卷，665，2021 650页 – 。[32] R. Luss、P. Chen、A. Dhurandhar、P. Sattigeri、K. Shanmugam 和 C. Tu。利用单调属性函数生成对比解释。CoRR，2019。网址：http://arxiv.org/abs/1905.12698。[33] J. Marques-Silva、T. Gerspacher、M. C. Cooper、A. Ignatiev 和 N. Narodytska。单调分类器的解释。载于 M. Meila 和 T. Zhang 编，ICML 2021，机器学习研究论文集第 139 卷，第 7469–7479 页。机器学习研究会论文集，2021。[34] T. Miller。人工智能中的解释：来自社会科学的见解。人工智能，267:1–38，2019年。[35] C. Molnar。《可解释机器学习》。Lulu.com，2020年。ISBN 9780244768522。网址：https://books.google.fr/books?id= RHjTxgEACAAJ。[36] R. K. Mothilal、A. Sharma 和 C. Tan。通过多样化反事实解释来解释机器学习分类器。载于2020年公平、问责与透明度会议论文集，FAT* '20，第 607–617 页。计算机协会，2020年。[37] E. Noorani、P. Dissanayake、F. Hamman 和 S. Dutta。利用熵风险度量为模型集成生成反事实解释。载于 S. Das、A. Nowé 和 Y. Vorobeychik 编，第二十四届自主智能体与多智能体系统国际会议论文集，AAMAS，第 1566–1575 页。国际自主智能体与多智能体系统基金会 / 美国计算机协会，2025。[38] S. Rodriguez、J. Thangarajah 和 M. Winikoff。多智能体系统中基于需求的可解释性。载于 S. Das、A. Nowé 和 Y. Vorobeychik 编，第二十四届自主智能体与多智能体系统国际会议论文集，AAMAS，第 2726–2728 页。国际自主智能体与多智能体系统基金会 / 美国计算机协会，2025。[39] B. Smyth 和 M. T. Keane。少而精的反事实：生成可解释、可信且多样化的反事实解释。载于 M. T. Keane 和 N. Wiratunga 编，ICCBR，计算机科学讲义第 13405 卷，第 18–32 页。Springer，2022。[40] A. Tsiourvas、W. Sun 和 G. Perakis。面向神经网络的流形对齐反事实解释。载于人工智能与统计国际会议，机器学习研究会论文集第 238 卷，第 3763–3771 页，2024。[41] S. Tsirtsis 和 M. G. Rodriguez。决策、反事实解释与策略行为。载于 H. Larochelle 、 M. Ranzato 、 R. Hadsell 、 M. Balcan 和 H. Lin 编，神经信息处理系统，2020年。

[42] S. Verma、J. P. Dickerson 和 K. Hines。机器学习的反事实解释：重新审视挑战。*CoRR*，abs/2106.07756，2021。网址：https://arxiv.org/abs/2106.07756。[43] S. Verma、K. Hines 和 J. P. Dickerson。黑盒模型的序列反事实解释的摊销生成。*CoRR*，abs/2106.03962，2021。网址：https://arxiv.org/abs/2106.03962。[44] S. Verma、V. Boonsanong、M. Hoang、K. Hines、J. Dickerson 和 C. Shah。机器学习的反事实解释与算法补救：综述。*ACM Computing Surveys*，56(12):312:1–312:42，2024。[45] J. von Kügelgen、A.-H. Karimi、U. Bhatt、I. Valera、A. Weller 和 B. Schölkopf。论因果算法补救的公平性，2022。[46] S. Wachter、B. D. Mittelstadt 和 C. Russell。无需打开黑盒的反事实解释：自动化决策与通用数据保护条例。*CoRR*，abs/1711.00399，2017。[47] G. Warren、B. Smyth 和 M. T. Keane。更好的反事实，让人们能够理解：用于可解释人工智能（XAI）的、心理上可信的基于案例的反事实，采用类别特征。载于基于案例的推理研究与发展——第三十届国际会议，*ICCBR*，计算机科学讲义第 13405 卷，第 63–78 页。Springer，2022。[48] G. Warren、E. Delaney、C. Guéret 和 M. T. Keane。以反事实方式解释多个实例：针对 XAI 群体反事实的用户测试。载于第三十二届国际基于案例推理研究与发展会议论文集，*ICCBR*，第 14775 卷，第 206–222 页，2024。[49] M. Winikoff、J. Thangarajah 和 S. Rodriguez。可解释人工智能评分表。载于第二十四届自主智能体与多智能体系统国际会议论文集，*AAMAS*，第 2171–2180 页，2025。[50] J. Woodward 和 C. Hitchcock。解释性概括，第一部分：一种反事实论述。*Noûs*，37(1):1–24，2003。

一个示例

我们提供一个新的示例，它对本文档中给出的部分证明很有用。

例 2. 考虑下文所述的理论以及分类器 κ_1的结果。

$\mathbb{F}(\mathbb{T}_1)$	f_1	f_2	$\kappa_1(x_i)$
x_1	*0*	*0*	c_1
x_2	*0*	*1*	c_2
x_3	*1*	*0*	c_2
x_4	*1*	*1*	c_3

核心(c_1)={$(f_1,\ 0)$，$(f_2,\ 0)$}

核心(c_2)=∅

核心(c_3)={$(f_1,\ 1)$，$(f_2,\ 1)$}

B 性质的证明

下面列出了一些用于证明主要定理的性质。

性质 1。令 $\mathbb{T} \in \mathtt{Th}$。 $\forall x \in \mathbb{F}(\mathbb{T})$， $\forall E \in \mathbb{E}(\mathbb{T})$， $x_{\downarrow E} \in \mathbb{F}(\mathbb{T})$。

证明。 由 $x_{\downarrow E}$的定义以及部分赋值的一致性可知，即对于任意部分赋值 E， $\nexists l,\ l' \in E$，满足 $l = (f,\ v)$、 $l' = (f',\ v')$，且 $f = f'$而 $v \neq v'$。 □

性质 2. 令 $\mathbb{T} = \langle \mathtt{F}、\mathtt{dom}、\mathtt{C} \rangle$ 为一个理论， $c \in \mathtt{C}$。

$$\mathtt{Core}(c) = \bigcap_{(x \in \mathbb{F}(\mathbb{T})) \wedge (\kappa(x) = c)} x.$$

证明。 由定义 1 可直接得出。 □

性质 3. 设$\mathbf{Q} = \langle \mathbb{T}, \kappa, x \rangle$ 为一个查询。 $\forall y \in \mathbb{F}(\mathbb{T})$ 使得$\kappa(y) \neq \kappa(x)$，$y \in \mathtt{gSuf}(\mathbf{Q})$。

证明。 由定义 6 可得。 □

性质 4。 设 $\mathbf{Q} = \langle \mathbb{T}, \kappa, x \rangle$ 为一个查询。

- cSuf(**Q**) $= \{E \in \mathbb{E}(\mathbb{T}) \mid x \cap E = \emptyset$ 和 $\kappa(x_{\downarrow E}) \neq \kappa(x)\}$，
- cSuf(**Q**) $\neq \emptyset$。

证明。 设$\mathbf{Q} = \langle \mathbb{T}, \kappa, x \rangle$ 是一个查询，且 $E \in \mathbb{E}(\mathbb{T})$。

• 假设 $E \in$ cSuf(**Q**)，因此存在 $E = y \setminus x$ 使得 $\kappa(y) \neq \kappa(x)$。因此，$E \cap x = \emptyset$ 且 $x_{\downarrow E} = y$。• 现在假设 $x \cap E = \emptyset$ 且 $\kappa(x_{\downarrow E}) \neq \kappa(x)$。令 $y = x_{\downarrow E}$。由此可得 $E = y \setminus x$， 因此 $E \in$ cSuf(**Q**)。• 我们证明cSuf(**Q**) $\neq \emptyset$。由于 κ 是满射且 $|\mathsf{C}| > 1$，我们有 $\exists y \in \mathbb{F}(\mathbb{T})$使得 $\kappa(y) \neq \kappa(x)$，因此 $y \setminus x \in$ cSuf(**Q**)。

□

C 命题证明

命题 1. 以下蕴含关系成立。

• 核心性 $\Rightarrow$ 可行性。• 核心性和非平凡性 $\Rightarrow$ 怀疑有效性。• 怀疑有效性 $\Rightarrow$ 非平凡性。• 弱有效性 $\Rightarrow$非平凡性。• 强有效性 $\Rightarrow$ 弱有效性。• 新颖性和非平凡性 $\Rightarrow$ 怀疑有效性。

证明。 设 **L** 是一个解释器，$\mathbf{Q} = \langle \mathbb{T}, \kappa, x \rangle$ 是一个查询，且 $E \in$**L**(**Q**)。

• 设 L 满足核心性。因此， $E \subseteq$ Core($\kappa(x)$). 根据性质 2， $E \subseteq \bigcap y$，所以 $E \subseteq x$，这表明 L $(_{y \in \mathbb{F}(\mathbb{T})}) \wedge (\kappa$ $(_y) = \kappa(x))$ 满足可行性。• 假设 L 满足核心性和非平凡性。根据非平凡性， $E \neq \emptyset$。根据核心性， $E \subseteq$ Core($\kappa(x)$)。根据性质 2， $E \subseteq \bigcap y$. 假设存在某个 $z \in \mathbb{F}(\mathbb{T})$ $(_{y \in \mathbb{F}(\mathbb{T})}) \wedge$ $(\kappa(y) = \kappa(x))$，使得 $x \setminus z = E$，那么 $E \not\subseteq z$。因此，$\bigcap y \not\subseteq z$， $(_{y \in \mathbb{F}(\mathbb{T})}) \wedge (\kappa(y) = \kappa(x))$，于是 $\kappa(z) \neq \kappa(x)$，这表明怀疑有效性成立。• 假设 L 满足怀疑有效性。那么， $\forall$Q $= \langle \mathbb{T}, \kappa, x \rangle$， $\forall E \in \forall y \in x \ominus E\ \kappa\ y \neq \kappa\ x \emptyset \notin$ L(Q), , ( ) ( )。显然，L(Q)，否则我们会得到 $\kappa(x) \neq \kappa(x)$。这表明 L 满足非平凡性。• 假设 L 满足弱有效性。那么， $\forall$Q $= \langle \mathbb{T}, \kappa,$ $x \rangle$, $\forall E \in$L(Q), $\kappa(x_{\downarrow E}) \neq \kappa(x)$。显然， $\emptyset \notin$L(Q)，否则我们会得到 $\kappa(x) \neq \kappa(x)$。因此 L 满足非平凡性。• 假设 L 满足强有效性。令 $E \in \mathbb{E}(\mathbb{T})$ 且 $y = x_{\downarrow E}$ $E \subseteq y$ $\kappa\ y \neq \kappa\ x$ 。因此，。根据强有效性，( ) ( )。所以，L 满足弱有效性。• 假设 L 满足新颖性和非平凡性。再假设 L 不满足怀疑有效性。那么， $\exists E \in$ L(Q)， $\exists y \in x \ominus E$使得 $\kappa(y) = \kappa(x)$。因此， $E \subseteq x$ (C)。根据非平凡性，$E \neq \emptyset$；根据新颖性， $E \cap x = \emptyset$，这与 (C) 矛盾。因此，L 满足怀疑有效性。

□

命题 2。成立的是gNec $\sqsubseteq$ sNec。

证明。 令$\mathbf{Q} = \langle \mathbb{T}, \kappa, x \rangle$ 为一个查询，使得 $\kappa(x) = c$。我们证明 gNec(**Q**) $\subseteq$sNec(**Q**)。令 $E \in$gNec(**Q**)。根据定理 3， $E \subseteq$ Core(c)。根据性质 2， $E \subseteq \bigcap y$。此外， E是非空的$_{\kappa(y) = \kappa(x)}$。对于任意 $y \in x \ominus E$、 $x \setminus y = E$，因此 $E \cap y = \emptyset$，于是 $\kappa(y) \neq \kappa(x)$，这意味着 $E \in$ sNec(**Q**)。 □

命题 3。令 $\mathbf{Q} = \langle \mathbb{T}, \kappa, x \rangle$ 为一个查询。

- sSuf $\sqsubseteq$gSuf。
- sSuf $\sqsubseteq$cSuf。
- 对于任意 $E \in \mathtt{cSuf}(\mathbf{Q})$，$\exists E' \in \mathtt{gSuf}(\mathbf{Q})$ 使得 $E = E' \setminus x$。
- 对于任意 $E \in \mathtt{gSuf}(\mathbf{Q})$，$E \setminus x \in \mathtt{cSuf}(\mathbf{Q})$。

证明。令 $\mathbf{Q} = \langle \mathbb{T}, \kappa, x \rangle$ 为一个查询。

- sSuf $\sqsubseteq$ gSuf 的包含关系显然。- 令 $E \in$ $E \cap x = \emptyset$ $\forall y \in \mathbb{F}\ \mathbb{T}$ $E \subseteq$ sSuf(**Q**)。因 $y\ \kappa\ y\ \neq \kappa\ x$ $E \in \mathbb{E}\ \mathbb{T}$ 1 $x_{\downarrow E} \in \mathbb{F}\ \mathbb{T}$ 此，且存在 使得，( ) ( )。由于 ( )，根据性质，( )。令 $z = x_{\downarrow E}$。根据 $E \cap x = \emptyset, E = z \setminus x$。由于 $E \subseteq z$，我们有 $\kappa(z) \neq \kappa(x)$。因此 $E \in$ cSuf(**Q**)。- 令 $E \in$ cSuf(**Q**)。因此，$\kappa(x_{\downarrow E}) \neq \kappa(x)$。令 $y = x_{\downarrow E}$。根据性质 1，$y \in \mathbb{F}(\mathbb{T})$。根据性质 3，$y \in$ gSuf(**Q**)。由于 $E \cap x = \emptyset$, $E = y \setminus x$。- 令 $E \in$ gSuf(**Q**)。令 $E' = E \setminus x$ 和 $y = x_{\downarrow E'}$。由于 $E \subseteq y$, κ x $(y) \neq \kappa(\ )$, 因此 $E' \in$ cSuf(**Q**)。

□

命题 4。令 $\mathbf{Q} = \langle \mathbb{T}, \kappa, x \rangle$ 为一个查询。

- 对于任意 $E \in$ gNec(**Q**)，– $E \subseteq x$– $E \subseteq$核心($\kappa(x)$)。
- 对于任意 $E \subseteq$ 核心($\kappa(x)$) </style}使得 、gNec(**Q**)。

证明。令 $\mathbf{Q} = \langle \mathbb{T}, \kappa, x \rangle$ 为一个问题。

- 假设 $E \in$ gNec(**Q**)
  - 根据定义 4，$E \not\subseteq x$ 将推出矛盾 $\kappa(x) \neq \kappa(x)$。– 根据定义 4，$\forall y \in \mathbb{F}(\mathbb{T})$ 使得 $E \not\subseteq y$、$\kappa(y) \neq \kappa(x)$。因此，$\forall l \in E$、$l \in$ 核心($\kappa(x)$)，从而 $E \subseteq$ 核心($\kappa(x)$)。

- 现在令 $E \subseteq$ 核心($\kappa(x)$). 现在假设 $E \neq \emptyset$。根据性质 2，$\forall y \in \mathbb{F}(\mathbb{T})$, 使得 $\kappa(y) = \kappa(x)$，(1)。现在假设 $E \notin$ gNec(**Q**)。因此，$\exists y \in \mathbb{F}(\mathbb{T})$, 使得 $E \not\subseteq y$ 且 $\kappa(y) = \kappa(x)$，这与 (1) 矛盾。因此 $E \in$ gNec(**Q**)。

□

命题5. 令 $\mathbf{Q} = \langle \mathbb{T}, \kappa, x \rangle$ 为一个查询。对于任意 $E \subseteq$ 核心($\kappa(x)$)，满足 $E \neq \emptyset, E \in \mathtt{sNec}(\mathbf{Q})$。

证明。令 $\mathbf{Q} = \langle \mathbb{T}, \kappa, x \rangle$ 为一个查询，$E \subseteq$ 核心($\kappa(x)$)，使得 $E \neq \emptyset$。假设 $E \notin$ 2 $E \subseteq x$ $\exists y \in$ sNec(**Q**)。根据性质，。 $\mathbb{F}\ \mathbb{T}$ 2 因此，( )，使得 $x \setminus y = E$ 且 κ $= \kappa\ x$ (y) ( )。根据性质，$E \subseteq y$ $x \setminus y = E$ $E \neq \emptyset$，若且则不可能。□

D 定理的证明

定理1. 每个集合($\mathtt{I}_i$) 的公理都是不相容的。

- ($\mathtt{I}_1$) 成功性、非平凡性和核心性。- ($\mathtt{I}_2$) 成功性、可行性和怀疑有效性。- ($\mathtt{I}_3$) 成功性、新颖性和强有效性。- ($\mathtt{I}_4$) 成功性、非平凡性、可行性和新颖性。- ($\mathtt{I}_5$) 成功性、可行性和弱有效性。- ($\mathtt{I}_6$) 成功性、非平凡性、等价性和可行性。- ($\mathtt{I}_7$) 成功性、非平凡性、等价性和新颖性。

证明。令 $\mathtt{I}_i$ 满足 $i \in \{1, ..., 7\}$。

$\mathtt{I}_1$ 设 L 是一个解释器，$\mathbf{Q} = \langle \mathbb{T}, \kappa, x \rangle$ 是一个查询，且 Core($\kappa(x)$) $= \emptyset$。假设 L 满足核心性和成功性。根据成功性，L(Q) $\neq \emptyset$。因此，$\exists E \in$ L(Q)。根据核心性，$E \subseteq$Core($\kappa(x)$)，所以 $E = \emptyset$，这意味着 L违反了非平凡性。$\mathtt{I}_2$ 考虑一个由两个特征 $a = \{0, 1\}$ $b = \{0, 1, 2\}$ κ, b 构成的理论，使得 dom( ) 和 dom( )。考虑如下二元分类器：当且仅当 $x = ((a, 0), (b, 1))$ 时，$\kappa(x) = 1$。考虑实例 $z = ((a, 0), (b, 0))$（其中 $\kappa(z) = 0$）及其对应的查询 Q。假设解释器 L 满足可行性、怀疑有效性和成功性。根据可行性，L(Q) $\subseteq \mathbb{P}(z)$（其中 $\mathbb{P}(z)$ 是 z 的幂集）。根据成功性，$\exists E \in$ L(Q)。容易验证，无论 E 的取值如何，$\exists y \in z \ominus E$ 都满足 $\kappa(y) = \kappa(z)$，这与满足怀疑有效性相矛盾。

$\mathtt{I}_3$ 假设**L** 满足成功性、新颖性和强有效性。考虑上文中的同一理论，但在 $\mathtt{I}_2$ 的情形下，这次考虑实例 $w = ((a, 1), (b, 1))$。根据成功性，**L(Q)** 非空。唯一满足强有效性的解释是 $\{(a,0),(b,1)\}$，但这个解释违反了新颖性。

$\mathtt{I}_4$ 假设**L** 满足成功性和非平凡性，那么 $\exists E \in$ **L(Q)**，使得 $E \neq \emptyset$。如果**L** 满足新颖性，那么 $x \cap E = \emptyset$，这意味着**L**违反了可行性。

$\mathtt{I}_5$ 假设L满足成功性和L(Q)，使得 $E \subseteq x$。因此，$x_{\downarrow E} = x$ 和 $\kappa(x_{\downarrow E}) = \kappa(x)$，这违反了弱有效性。

$\mathtt{I}_6$ 设 Q $= \langle \mathbb{T}, \kappa, x \rangle$ 和 Q′ $= \langle \mathbb{T}, \kappa, y \rangle$ 是满足 $\kappa(x) = \kappa(y)$ 和 $x \cap y = \emptyset$ 的查询。假设 L 满足成功性、非平凡性、等价性和可行性。根据成功性，L(Q) $\neq \emptyset$ 且 L(Q′) $\neq \emptyset$。根据非平凡性，$\emptyset \notin$ L(Q) 且 $\emptyset \notin$ L(Q′)。因此，$\exists E \in$ L(Q)、$\exists E' \in$ L(Q′)，使得 $E \neq \emptyset$ 和 $E' \neq \emptyset$。根据可行性，$E \subseteq x$ 和 $E' \subseteq y$。因此，$E \notin$ L(Q′) 且 $E' \notin$ L(Q)。所以，L(Q) $\neq$ L(Q′)，而等价性意味着 L(Q) $=$ L(Q′)。

$\mathtt{I}_7$ 假设 L 满足成功性、非平凡性、等价性和新颖性。设 Q $= \langle \mathbb{T}, \kappa, x \rangle$ 和 Q′ $= \langle \mathbb{T}, \kappa, y \rangle$ 为查询，其中$\mathbb{T}$ 由两个二元特征 f_1 和 f_2组成，且 κ 是一个二元分类器，满足 $\kappa(x) = f_1 \equiv f_2$。设 $x = \langle (f_1, 0), (f_2, 0) \rangle$ 以及 $y = \langle (f_1, 1), (f_2, 1) \rangle$。那么，$\kappa(x) = \kappa(y) = 1$ 且 $x \cap y = \emptyset$。根据等价性，L(Q) $=$ L(Q′)。根据成功性，$\exists E \in$ L(Q)，因此 $E \in$ L(Q′)。根据新颖性，$E \cap x = \emptyset$ 且 $E \cap y = \emptyset$，由此可推出 $E = \emptyset$（因为 x 和 y 涵盖了 $\mathbb{T}$的两个变量的所有可能赋值）。但这样一来，非平凡性就无法成立。

□

定理 2. 命题*1*和定理*1*未排除的所有公理组合，都能由某个解释器满足。

证明。我们考虑解释器 cSuf 和 gSuf，以及以下三个解释器：

- $\mathbf{L}_0$ 定义为 $\forall \mathbf{Q}, \mathbf{L}_0(\mathbf{Q}) = \emptyset$. - $\mathbf{L}_1$ 定义为 $\forall \mathbf{Q}, \mathbf{L}_1(\mathbf{Q}) = \{\emptyset\}$.
- $\mathbf{L}_2$ $\forall$ $= \langle \mathbb{T}\ \kappa\ x \rangle$ $_2$ $= \{x \setminus y \mid \kappa\ y \neq$定义为 $\mathbf{Q}$,,, $\mathbf{D}$, κ, x $\mathbf{L}(\mathbf{Q})$ ( ) ( )}.

显然，$\mathbf{L}_0$ 和$\mathbf{L}_1$ 并不是有用的解释器，因为它们是常函数，实际上不会提供任何信息。解释器$\mathbf{L}_2$可以看作是信息量较少的 cSuf：对于每个满足 $\kappa(y) \neq \kappa(x)$ 的 $y \in \mathbb{F}$，cSuf 会返回 $y \setminus x$，即需要更改的

特征的新值，而$\mathbf{L}_2$返回信息量较少的 $x \setminus y$，即需要更改的特征的旧值。

$\mathbf{L}_0$ 显然满足除成功性之外的所有公理。因此，我们只需考虑包含成功性的公理集合。

根据命题 1，如果某个解释器不满足非平凡性，那么它就不能满足怀疑有效性、弱有效性或强有效性。很容易验证，$\mathbf{L}_1$ 满足其他五条公理。因此，我们只需考虑同时包含成功性和非平凡性的公理集合。

如果某个解释器既不满足可行性，也不满足新颖性，那么它就不能满足核心性（根据命题 1）。从表 2 可以看出，解释器 gSuf 满足其他六条公理。因此，我们只需考虑满足可行性或新颖性（以及成功性和非平凡性）的公理集合。

如果某个解释器满足成功性、非平凡性和可行性，那么根据定理1和命题 1，它就不能满足其他六条公理。成功性、非平凡性和可行性这几条公理彼此相容，因为$\mathbf{L}_2$满足它们。

最后，如果一个解释器满足成功性、非平凡性和新颖性，那么根据命题 1 和定理1，它不可能满足核心性、可行性、等价性或强有效性。由于 cSuf 满足其余五条公理（见表2），这就完成了证明。 □

定理 3. 解释器 $\mathbf{L}$ 满足核心性和非平凡性当且仅当 $\mathbf{L} \sqsubseteq \mathtt{gNec}$。

Proof. Let $\mathbf{Q} = \langle \mathbb{T}, \kappa, x \rangle$ be a query.

• 设 L 是一个满足核心性和非平凡性的解释器。因此，对于任意 $E \in \mathrm{L}(\mathrm{Q})$，$E \subseteq \mathrm{Core}(\kappa(x))$ 和 $E \neq \emptyset$。根据性质 2，$E \subseteq \bigcap_{y \in \mathbb{F}(\mathbb{T}) \text{ s.t. } E \not\subseteq (y \in \mathbb{F}(\mathbb{T})) \wedge (\kappa(y)=\kappa(x))} y$。因此，$\forall y \in \mathbb{F}(\mathbb{T})$ s.t. $E \not\subseteq (y \in \mathbb{F}(\mathbb{T})) \wedge (\kappa(y)=\kappa(x))$ y, $\kappa(y) \neq \kappa(x)$。由此，$E \in \mathrm{gNec}(\mathrm{Q})$。• 现在假设有一个解释器 L，使得 $\mathrm{L} \sqsubseteq \mathrm{gNec}$。如果对于任意查询 Q，$\mathrm{L}(\mathrm{Q}) = \emptyset$，那么 L 显然满足核心性和非平凡性。现在假设 $E \in \mathrm{L}(\mathrm{Q})$。那么，$E \in \mathrm{gNec}(\mathrm{Q})$，因此 $E \neq \emptyset$。根据命题 4 的第一项，$E \subseteq \mathrm{Core}(\kappa(x))$。因此，L 满足核心性和非平凡性。

□

定理4。 表*2*中的性质成立。

证明。 设$\mathbf{L}$ 为解释器，$\mathbf{Q} = \langle \mathbb{T}, \kappa, x \rangle$ 为一个查询。

函数 $\mathbf{L} \sqsubseteq \mathtt{gNec}$。

• 非平凡性和核心性。由定理 3 得出。• 成功性。根据定理1，由于 L 满足核心性和非平凡性，因此违反成功性。• 等价性。gNec 满足这一公理，但函数 $L \sqsubseteq \mathrm{gNec}$ 可能存在这样的情况：对于两个实例 x、y，使得 $\kappa(x) = \kappa(y)$，$\mathrm{L}(x) \neq \mathrm{L}(y)$。• 可行性。根据命题 4，$\forall E \in \mathrm{L}(\mathrm{Q})$，$E \subseteq x$，因此满足可行性。• 怀疑有效性。根据核心性和非平凡性可得（依据命题 1）。• 新颖性。对于任意查询 Q，若解释器 L 满足 $= \emptyset$，则满足新颖性。然而，gNec 违反该公理。• 强有效性和弱有效性。对于任意查询 Q，若解释器 L 满足 $= \emptyset$，则显然满足强有效性和弱有效性。然而，gNec 违反这两条公理。

函数 $\mathbf{L} = \mathtt{gNec}$。

• 核心性和非平凡性。 由定理 3 得出。• 成功性。 根据定理1，由于 $\mathbf{L}$ 满足核心性和非平凡性，因此违反成功性。• 等价性。 由命题 4 得出，即对于任意 $\mathbf{Q}$，$\mathrm{gNec}(\mathbf{Q}) = \{E \subseteq \mathrm{Core}(\kappa(x)) \mid E \neq \emptyset\}$。• 可行性和怀疑有效性。 由任何满足可行性和怀疑有效性的 $\mathbf{L} \sqsubseteq \mathrm{gNec}$ 这一事实得出。• 新颖性和弱有效性。 考虑查询 $\mathbf{Q} = \langle \mathbb{T}, \kappa, x \rangle$ 使得 $\mathrm{Core}(\kappa(x)) \neq \emptyset$。那么 $\exists E \in \mathrm{gNec}(\mathbf{Q})$ 使得 $E \neq \emptyset$ 且 $E \subseteq \mathrm{Core}(\kappa(x))$。根据性质 2，$E \subseteq x$。

– 由于 $x \neq \emptyset$，因此 $x \cap E \neq \emptyset$，这与新颖性矛盾 .

– 由于 $E \subseteq x$ $x_{\downarrow E} = x$ $\kappa\ x_{\downarrow E} \neq$，因此 。

根据弱有效性，( ) ( )，这是不可能的。

• 强有效性 gNec 违反弱有效性，因此根据命题 1，即强有效性 ⇒ 弱有效性这一蕴含关系，gNec 违反强有效性。

函数 $\mathbf{L} \sqsubseteq \mathtt{sNec}$。

• 可行性和怀疑有效性。 根据定理 5。• 非平凡性。 根据命题 1，非平凡性由怀疑有效性推出。• 成功性。 根据定理1，成功性、怀疑有效性和可行性这三条公理互不相容。因此，$\mathbf{L}$ 违反成功性。• 等价性、核心性、新颖性、强有效性、弱有效性。 解释器$\mathbf{L}$ 使得对于每个查询$\mathbf{Q}$，$\mathbf{L}(\mathbf{Q}) = \emptyset$ 满足所有这些公理，而 sNec 违反它们。

函数 $\mathbf{L} = \mathtt{sNec}$。

• 成功性。这源于任何$\mathbf{L} \sqsubseteq$ sNec 都违反该公理这一事实。

• 可行性、非平凡性、怀疑有效性。由任何满足这些公理的 L $\sqsubseteq$sNec 这一事实可知。• 等价性。考虑一个由两个二元特征 f_1、f_2 和二元分类器 κ 构成的理论，使得 $\kappa(x) = f_1 \vee f_2$。令 $x = ((f_1,0),(f_2,1))$，且令 $y = ((f_1,1),(f_2,0))$。显然，$\kappa(x) = \kappa(y)$。注意，$\{((f_2,1))\}$ 是 $\kappa(x)$ 的一个怀疑式必要理由，而它不是 $\kappa(y)$ 的理由，因为 $(f_2,1) \notin y$。• 核心性、新颖性、强/弱有效性。考虑下面的理论和分类器。

示例 3. 令$\mathbf{Q}_1$ 为关于 x_1的查询。

$\mathbb{F}(\mathbb{T}_2)$	f_1	f_2	$\kappa_2(x_i)$
x_1	0	0	c_1
x_2	0	1	c_1
x_3	1	0	c_2
x_4	1	1	c_1

E_1=∅
E_2={(f_1, 0)}
E_3={(f_2, 0)}
E_4=x_1

注意，$\mathtt{sNec}(\mathbf{Q}_1) = \{E_2\}$而$(f_1,\ 0) \notin$核心$(c_1)$，这违反了核心性。此外，$E_2 \cap x_1 \neq \emptyset$这违反了新颖性，$E_2 \subseteq x_2$ 而 $\kappa(x_1) = \kappa(x_2)$ 这违反了强有效性，最后，$\kappa(x_{1 \downarrow E_2}) = \kappa(x_1)$ 这违反了弱有效性。

函数 $\mathbf{L} \sqsubseteq \mathtt{gSuf}$。

• 成功性。 函数$\mathbf{L}$ 违反该公理：对于每个查询$\mathbf{Q}$，$\mathbf{L}(\mathbf{Q}) = \emptyset$；而 gSuf 满足该公理。

• 非平凡性。 假设

$$\emptyset \in \qquad \forall y \in \mathbb{F}\ \mathbb{T}\quad \kappa\ y\ \neq \mathbf{L(Q)}。$$

$\kappa\ x$ 因此，( )，( ) ( )，这不可能，因为 x。所以，**L** 满足该公理。• 可行性、核心性、新颖性。 函数**L** 满足这些公理：对于每个查询**Q**，**L(Q)** $= \emptyset$；而 gSuf 违反这些公理。• 怀疑有效性。 分两种情况：– 对于每个查询 **Q**，**L(Q)** $= \emptyset$。**L**显然满足该公理。– 对于某个查询**Q**，**L(Q)** $\neq \emptyset$。根据非平凡性，$\emptyset \notin$ **L(Q)**。因此，$\exists E \in$ **L(Q)**，$E \neq \emptyset$。此外，$E \not\subseteq x$ (**C**)，否则根据定义

$$8\ \kappa\ x\ \neq \kappa\ x \qquad x \ominus E = \{y \in, (\ )\ (\ )。回$$

$\mathbb{F}\ \mathbb{T}\ \mid x \setminus y = E\}\quad x \ominus E = \emptyset$ 想一下，( )，因此（由(**C**)）。于是，**L**以一种平凡的方式满足怀疑有效性。

• 强有效性。 由定理 6 得出。• 弱有效性。 根据命题 1，由强有效性得出。

函数 L $=$ gSuf。

• 成功性。 由性质 3 以及 κ 是满射且一个理论包含多个类别这一事实得出。• 非平凡性。 这是因为每个函数 **L** $\sqsubseteq$ gSuf 都满足公理。• 等价性。由 gSuf 的定义得出。• 可行性。 由于 gSuf 满足成功性，因此对于每个查询 **Q**，gSuf(**Q**) $\neq \emptyset$。令 $E \in$ gSuf(**Q**)。假设 $E \subseteq x$。根据定义 8，我们得到矛盾 κ(x) $\neq \kappa(x)$。• 核心性。 这是因为 gSuf 满足成功性和非平凡性，而二者与核心性不相容（见定理 1）。

• 怀疑有效性。 这可由以下事实推出：每个 **L** $\sqsubseteq$ gSuf 都满足此公理。

• 新颖性。 根据示例 2，$x_2 \in$ gSuf(**Q**$_1$)，而 $x_2 \cap x_1 \neq \emptyset$。 • 强有效性。 根据定理 6。 • 弱有效性。 由强有效性和命题 1 得出。

函数 L $\sqsubseteq$ sSuf、sSuf、**L** $\sqsubseteq$ cSuf **以及**cSuf。

• 成功性。 考虑关于实例 x_2 的查询**Q**$_2$，见示例 2。注意，sSuf(**Q**$_2$) $= \emptyset$，因此 sSuf 以及任何**L** $\sqsubseteq$ sSuf 都违反成功性。现在我们证明 cSuf 满足成功性。由于 $|\mathbb{C}| > 1$ 且 κ 是满射，因此 $\exists y \in \mathbb{F}(\mathbb{T})$，使得 $\kappa(y) \neq \kappa(x)$，所以 $y \setminus x$ 是针对 $\kappa(x)$ 的一个确信型理由。令**L** $\sqsubseteq$ cSuf。若**L** 对每个查询都返回空集，则它违反成功性。然而，若**L** $=$ cSuf，则它满足此公理。

• 非平凡性。注意，$\emptyset \notin$ cSuf(Q)，因为 $\kappa(x_{\downarrow\emptyset}) = \kappa(x)$。因此，L $\sqsubseteq$ cSuf，且 cSuf 满足非平凡性。根据命题 3，sSuf $\sqsubseteq$ cSuf。因此，sSuf 满足非平凡性。由此，每个 L $\sqsubseteq$ sSuf 都满足该公理。• 等价性。考虑示例 2，并令 $\mathsf{Q}_i = \langle \mathbb{T}, \kappa, x_i \rangle$，其中 $i = 1, \ldots, 4$。注意，$\kappa(x_2) = \kappa(x_3)$，而 cSuf(Q$_3$ = {{ f_1 0 } { f_2 1 }})

$= \{\{(f_1, 1)\}, \{(f_2, 0)\}\}$，而 cSuf(Q) (,) , (,)。这表明 cSuf 违反等价性。在示例 2 中，$\kappa(x_1) = \kappa(x_4)$，但 sSuf(Q1) $\neq$ sSuf(Q4)。这表明 sSuf 违反等价性。对于 L $\sqsubseteq z$，当 $z \in \{$ sSuf、cSuf$\}$ 时，若 $\forall$Q，L(Q) $= \emptyset$，则 L 显然满足等价性；但当 L $= z$ 时，L 违反等价性。

• 可行性。cSuf 违反该公理，因为它满足成功性和弱有效性，而这两者与可行性不相容（见定理1）。sSuf 违反该公理，因为它满足成功性和强有效性，而这两者与可行性不相容（见定理1）。对于 L $\sqsubseteq z$，当 $z \in \{$sSuf、cSuf$\}$ 时，若 $\forall$Q，L(Q) $= \emptyset$，则 L 满足该公理。当 L $= z$ 时，L 违反该公理。• 核心性。cSuf 和 sSuf 都违反核心性，因为二者均违反可行性；根据命题 1，我们有如下蕴含关系：核心性 $\Rightarrow$可行性。对于 L $\sqsubseteq z$，当 $z \in \{$sSuf、cSuf$\}$ 时，若 $\forall$Q，L(Q) $= \emptyset$，则 L 满足该公理。当 L $= z$ 时，L 违反该公理。• 怀疑有效性。令 L $\sqsubseteq z$，其中 $z \in \{$sSuf、cSuf$\}$。若 $\forall$Q，L(Q) $= \emptyset$，则 L 以显然的方式满足该公理。现在假设 $E \in$L(Q)。根据新颖性，$E \cap x = \emptyset$。因此，$x \ominus E = \emptyset$，这意味着 L 以显然的方式满足该公理。由此可知，sSuf 和 cSuf 满足该公理。• 新颖性。根据定理7和定理 8。• 强有效性。对于 L $\sqsubseteq$ sSuf 和 sSuf，该性质由定理7得出。正如示例 1 中 Q3 的 CSR $\{(a,$ 滑雪$)\}$ 所示，cSuf 违反该公理。函数 L $\sqsubseteq$ cSuf，满足 $\forall$Q，L(Q) $= \emptyset$，则满足该公理。

• 弱有效性。对于**L** $\sqsubseteq$ sSuf 和 sSuf，该性质可由满足强有效性和命题 1 得出。对于 **L** $\sqsubseteq$ cSuf 和 cSuf，该性质可由定理 8 得出。

□

定理 5。 解释器 **L** 满足可行性和怀疑有效性当且仅当**L** $\sqsubseteq$ sNec。

证明。 设**L** 是一个解释器。

• 假设 **L** 满足可行性和怀疑有效性。根据定义 5，可得 **L** $\sqsubseteq$ sNec。

• 假设**L** $\sqsubseteq$ sNec。根据定义5，对于所有 $E \in$ **L(Q)**，我们有：(1) $E \subseteq x$，因此**L** 满足可行性；(2) $\forall y \in x \ominus E, \kappa \qquad x \qquad (y) \neq \kappa(\ )$，因此**L**满足怀疑有效性。

□

定理 6。 解释器 **L** 满足强有效性当且仅当 **L** $\sqsubseteq$ gSuf。

证明。 设**L** 为解释器，且**Q** $= \langle \mathbb{T},\ \kappa,\ x \rangle$ 是一个查询。

• 假设**L**满足强有效性。根据强有效性的定义，对于任意 $E \in$ **L(Q)**，$\forall y \in \mathbb{F}(\mathbb{T})$ 使得 $E \subseteq y$，

$\kappa \qquad x \qquad (y) \neq \kappa(\ )$。因此，$E \in$ gSuf(**Q**)。所以，**L** $\sqsubseteq$ gSuf。• 假设**L** $\sqsubseteq$ gSuf。设 $E \in$ **L(Q)**。根据定义 6；$\forall y \in \mathbb{F}(\mathbb{T})$ 使得 $E \subseteq y, \kappa(y) \neq \kappa(x)$，因此**L**满足强有效性。

□

定理7。 解释器**L**满足新颖性和强有效性，当且仅当**L** $\sqsubseteq$ sSuf。

证明。 设**L** 为解释器，且**Q** $= \langle \mathbb{T},\ \kappa,\ x \rangle$ 是一个查询。

• 假设**L** 满足新颖性和强有效性。设 $E \in$ **L(Q)**。根据新颖性，

$$E \cap x = \emptyset \qquad \forall y \in 。根据强$$

$\mathbb{F}\ \mathbb{T} \qquad E \subseteq y\ \kappa\ y\ \neq \kappa\ x \qquad E \in$ 有效性，存在 ( ) 使得 , ( ) ( )。因此，sSuf(**Q**)，所以**L** $\sqsubseteq$ sSuf。

• 假设$\mathbf{L} \sqsubseteq$ sSuf。对于任意 $E \in \mathbf{L}(\mathbf{Q})$，根据定义 7：(1) $E \cap x = \emptyset$ 以及 (2) $\forall y \in \mathbb{F}(\mathbb{T})$ 使得 $E \subseteq y, \kappa(y) \neq \kappa(x)$。因此，$\mathbf{L}$ 满足这两条公理。

□

定理 8. 解释器 $\mathbf{L}$ 满足新颖性和弱有效性，当且仅当 $\mathbf{L} \sqsubseteq$ cSuf。

证明。 设 $\mathbf{L}$ 是一个解释器，且 $\mathbf{Q} = \langle \mathbb{T}, \kappa, x \rangle$ 是一个查询。

• 假设 $\mathbf{L}$ 满足新颖性和弱有效性。令 $E \in \mathbf{L}(\mathbf{Q})$。由新颖性，$E \cap x = \emptyset \quad \kappa(x_{\downarrow E}) \neq \kappa(x)$。由弱有效性，$\mathbf{L}(\mathbf{Q}) \sqsubseteq \mathrm{cSuf}(\mathbf{Q})$。因此，$\mathbf{L} \sqsubseteq$ cSuf。• 假设 $\mathbf{L} \sqsubseteq$ cSuf。令 $\mathbf{L}(\mathbf{Q})$。根据定义，$y \setminus x \in \mathbb{F}(\mathbb{T})$ 其中 $E \in \mathbb{E}(\mathbb{T})$。因此，$E \cap x = \emptyset$，这表明新颖性。注意 $y = E \cup (x \cap y)$。因此，$y = x_{\downarrow E}$。所以，根据定义 8，$\kappa(x_{\downarrow E}) \neq \kappa(x)$，这保证了弱有效性。

□

定理 9。 解释器 $\mathbf{L}$ 满足成功性、新颖性和弱有效性，当且仅当存在一个忠实排序 $\mathbf{S}$，将每个查询 $\mathbf{Q} = \langle \mathbb{T}, \kappa, x \rangle$ 映射为 $\mathbb{E}(\mathbb{T})$ 上的一个预序 $\succeq_{\mathbf{Q}}$，使得：

$$\mathbf{L}(\mathbf{Q}) = \max(\mathbb{E}(\mathbb{T}), \succeq_{\mathbf{Q}}).$$

证明。 设 $\mathbf{Q} = \langle \mathbb{T}, \kappa, x \rangle$ 是一个查询，$\mathbf{L}$ 是一个解释器。

令 $\mathbb{E}(\mathbb{T}) = \mathbb{E}_1(\mathbb{T}) \cup \mathbb{E}_2(\mathbb{T})$，其中

- $\mathbb{E}_1(\mathbb{T}) = \mathrm{cSuf}(\mathbf{Q}) = \{E \mid E \cap x = \emptyset$ 且 $\kappa(x_{\downarrow E}) \neq \kappa(x)\}$，
- $\mathbb{E}_2(\mathbb{T}) = \{E \mid E \cap x \neq \emptyset \quad \kappa(x_{\downarrow E}) = \kappa(x)\}$ 或 $\mathbb{E}_2(\mathbb{T}) = \emptyset$。

因此，$\mathbb{E}_1(\mathbb{T}) \cap \mathbb{E}_2(\mathbb{T}) = \emptyset$。

• 假设 $\mathbf{L}$ 满足这三个公理，并令 $E \in \mathbf{L}(\mathbf{Q})$（由成功性可知其存在）。根据解释器的定义，$E \in \mathbb{E}(\mathbb{T})$ (1)；新颖性蕴含 $E \cap x = \emptyset$ (2)；弱有效性蕴含 $\kappa(x_{\downarrow E}) \neq \kappa(x)$ (3)。因此，$\mathbf{L}(\mathbf{Q}) \subseteq \mathbb{E}_1(\mathbb{T})$。现在考虑一个函数$\mathbf{S}$，它将每个查询 $\mathbf{Q}$ 映射为定义在 $\mathbb{E}(\mathbb{T})$ 上的一个偏预序 $\mathbf{Q}$，且满足以下条件：

$$\forall E, E' \in \mathbb{E}(\mathbb{T}), \big(E \in \mathbb{E}_1(\mathbb{T})\big) \wedge \big(E' \in \mathbb{E}_2(\mathbb{T})\big) \Rightarrow E \succ_{\mathbf{Q}} E'.$$

并且使得 $\mathbf{L}(\mathbf{Q}) = \max(\mathbb{E}(\mathbb{T}) \succeq_{\mathbf{Q}})$。这是可行的，因为 $\mathbf{L}(\mathbf{Q}) \subseteq \mathbb{E}_1(\mathbb{T})$ 且 $\mathbf{L}(\mathbf{Q}) \neq \emptyset$。由于上述性质，函数 $\mathbf{S}$ 是一个忠实排序。

• 假设存在一个忠实排序 $\mathbf{S}$，它将任意查询 $\mathbf{Q}$ 映射为 $\mathbb{E}(\mathbb{T})$ 上的预序 $\succeq_{\mathbf{Q}}$，并且使得 $\mathbf{L}(\mathbf{Q}) = \max(\mathbb{E}(\mathbb{T}) \succeq)$。显然，$\mathbf{L}$满足成功性（因为非空集合总是至少有一个极大元素）。根据忠实排序的定义，$\succeq_{\mathbf{Q}}$ 满足以下条件。

$$\forall E, E' \in \mathbb{E}(\mathbb{T}), \big(E \in \mathbb{E}_1(\mathbb{T})\big) \wedge \big(E' \in \mathbb{E}_2(\mathbb{T})\big) \Rightarrow E \succ_{\mathbf{Q}} E'.$$

由于 $\mathbb{E}_1(\mathbb{T}) \neq \emptyset$ $\kappa \subseteq \mathbb{E}(\mathbb{T})$（根据 的满射性），可得 $\mathbf{L}(\mathbf{Q}) \mathbb{E}(\mathbb{T})$。因此，$\mathbf{L}$ 满足新颖性和弱有效性。

□

定理10. 以下性质成立：

• 将每个查询 $\mathbf{S}$ $\mathbf{Q}$ 映射到 $\succeq^f_{\mathbf{Q}}$ 在 $\mathbb{E}(\mathbb{T})$ 上的函数是一个忠实排序。

- *For any query* $\mathbf{Q}$, $\mathbf{L_{wf}}(\mathbf{Q}) = \max(\mathbb{E}(\mathbb{T}), \succeq^f_{\mathbf{Q}})$.

证明。 设 $\mathbf{Q}$ 为一个查询。根据定义 11，$\succeq^f_{\mathbf{Q}}$ 是一个偏预序。

• 设 $E, E' \in \mathbb{E}(\mathbb{T})$ 满足：

- $E \cap x = \emptyset$ 且 $\kappa(x_{\downarrow E}) \neq \kappa(x)$ 且
- $E' \cap x \neq \emptyset$ 或 $\kappa(x_{\downarrow E'}) = \kappa(x)$。

那么，$\delta(E) = 1$ 且 $\delta(E') = +\infty$，因此 $E \succ^f_{\mathbf{Q}} E'$，这意味着

$$\forall E \in \mathtt{cSuf}(\mathbf{Q}), \forall E' \notin \mathtt{cSuf}(\mathbf{Q}), E \succ^f E'. \qquad (1)$$

这表明函数 $\mathbf{S}$ 将每个查询 $\mathbf{Q}$ 映射到 $\mathbb{E}(\mathbb{T})$ 上的预序 $\succeq^f_{\mathbf{Q}}$，具有忠实性。

• 我们证明等式 $\mathbf{Q}, \mathbf{L_{wf}}(\mathbf{Q}) = \max(\mathbb{E}(\mathbb{T}) \succeq^f_{\mathbf{Q}})$。
$\max(\mathbb{E}(\mathbb{T}) \succeq^f \mathrm{Q}) = \{E \in \mathbb{E}(\mathbb{T}) \mid \nexists E' \in \mathbb{E}(\mathbb{T})\ s.t.\ E' \succ^f \mathrm{Q}E\}$。由 (1) 可得 $\max(\mathbb{E}(\mathbb{T}) \succeq^f \mathrm{Q}) \subseteq \mathrm{cSuf(Q)}$。因此，$\max(\mathbb{E}(\mathbb{T}) \succeq^f \mathrm{Q}) = \{E \in \mathrm{cSuf(Q)} \mid \nexists E' \in \mathrm{cSuf(Q)}\ s.t.\ E' \succ^f \mathrm{Q}\ E\}$。由于 $\forall E, E' \in \mathrm{cSuf(Q)}$，$\delta(E) = \delta(E')$，$\max(\mathbb{E}(\mathbb{T}) \succeq^f \mathrm{Q}) = \{E \in \mathrm{cSuf(Q)} \mid \nexists E' \in \mathrm{cSuf(Q)}\ s.t. \mathrm{Feat}(E') \subseteq \mathrm{Feat}(E)\} = \mathrm{L_{wf}(Q)}$。

□

定理 11. 以下性质成立：

• 将每个查询 $\mathbf{Q}$ 映射到 $\succeq^c_{\mathbf{Q}}$ 上的函数 $\mathbf{S}$ 在 $\mathbb{E}(\mathbb{T})$ 上是忠实的排序。• 对于任意查询 $\mathbf{Q}$，$\mathbf{L}_c(\mathbf{Q}) = \max(\mathbb{E}(\mathbb{T}) \succeq^c_{\mathbf{Q}})$。

证明。 根据定义13，$\succeq^c_{\mathbf{Q}}$ 是一个全预序。现在令$\mathbf{Q}$ 为一个查询。

• 令 $E, E' \in \mathbb{E}(\mathbb{T})$ 满足：

- $E \cap x = \emptyset$ and $\kappa(x_{\downarrow E}) \neq \kappa(x)$ and
- $E' \cap x \neq \emptyset$ or $\kappa(x_{\downarrow E'}) = \kappa(x)$.

那么，$\sigma(E) = |E|$且 $\sigma(E') = +\infty$，因此 $E \succ^c E'$ (由于特征集是有限的，$|E| < +\infty$)，这表明将每个查询 $\mathbf{Q}$ 映射到 $\succeq^c_{\mathbf{Q}}$ 上的函数$\mathbf{S}$在 $\mathbb{E}(\mathbb{T})$ 上具有忠实性。

• 我们证明等式 $\mathbf{Q}$、$\mathbf{L}_c(\mathbf{Q}) = \max(\mathbb{E}(\mathbb{T}) \succeq^c_{\mathbf{Q}})$。
$\max(\mathbb{E}(\mathbb{T}) \succeq^c_{\mathbf{Q}}) = \{E \in \mathbb{E}(\mathbb{T}) \mid \nexists E' \in \mathbb{E}(\mathbb{T})\ s.t.\ E' \succ^c E\} = \{E \in \mathbb{E}(\mathbb{T}) \mid \forall E' \in \mathbb{E}(\mathbb{T})\ E \succeq^c E'\}$（因为 $\succeq^c$ 是一个全预序）。所以，

$$\max(\mathbb{E}(\mathbb{T}), \succeq^c_{\mathbf{Q}}) = \{E \in \mathbb{E}(\mathbb{T}) \mid \forall E', \sigma(E) \leq \sigma(E')\}.$$

由于 $\forall E \in \mathrm{cSuf(Q)}$，$\sigma(E) = |E|$ 且 $\forall E' \notin \mathrm{cSuf(Q)}$，$\sigma E' = +\infty$ $\forall E \in \mathbb{E}(\mathbb{T}) \succeq^c_{\mathbf{Q}} \subseteq$ $\forall E' \notin \mathbb{E}(\mathbb{T}) \succeq^c_{\mathbf{Q}}$ $E \succ^c$，我们有 E' cSuf(Q)，cSuf(Q)，。因此，$\max(\mathbb{E}(\mathbb{T}))$ cSuf(Q)。所以，$\max(\mathbb{E}(\mathbb{T})) = \{E \in \mathrm{cSuf(Q)} \mid \nexists E' \in \mathrm{cSuf(Q)}\ s.t.\ E' \succ^c E\} = \mathbf{L}_c(\mathbf{Q})$。

□

定理12。 令$\mathtt{d}$ 为距离度量。

• 将每个查询$\mathbf{S}$ 映射到$\succeq^d_{\mathbf{Q}\ \mathtt{d}}$ 关于 $\mathbb{E}(\mathbb{T})$ 的函数是一个忠实排序。
• 对于任意查询$\mathbf{Q}$，$\mathbf{L}_{\mathtt{d}}(\mathbf{Q}) = \max(\mathbb{E}(\mathbb{T}) \succeq^d_{\mathbf{Q}\ \mathtt{d}})$。

证明。 设$\mathbf{Q}$是一个查询，d 是 $\mathbb{F}(\mathbb{T})$上的距离度量。因此，

$$\mathsf{d} : \mathbb{F}(\mathbb{T})^2 \to [0, k],$$

其中 $k < +\infty$。根据定义 15，$\succeq^d_{\mathbf{Q},\,\mathsf{d}}$ 是一个全预序。

- 令 E， $E' \in \mathbb{E}(\mathbb{T})$ 满足：
  - $E \cap x = \emptyset$ and $\kappa(x_{\downarrow E}) \neq \kappa(x)$ and
  - $E' \cap x \neq \emptyset$ or $\kappa(x_{\downarrow E'}) = \kappa(x)$.

  那么， $\gamma(E) = \mathsf{d}(x_{\downarrow E},\ x)$ 且 $\sigma(E') = +\infty$，因此 $E \succ^d_{\mathbf{Q},\,\mathsf{d}} E'$，这表明将每个查询 $\mathbf{Q}$映射到 $\succeq^d_{\mathbf{Q},\,\mathsf{d}}$ 的函数 $\mathbf{S}$ 具有忠实性，定义在 $\mathbb{E}(\mathbb{T})$ 上。
- 我们证明等式 $\mathbf{Q}, \mathbf{L_c(Q)} = \max(\mathbb{E}(\mathbb{T}) \succeq^d_{\mathbf{Q},\,\mathsf{d}})$.
  $\max(\mathbb{E}(\mathbb{T}) \succeq^d_{\mathbf{Q},\,\mathsf{d}}) = \{E \in \mathbb{E}(\mathbb{T}) \mid \nexists E' \in \mathbb{E}(\mathbb{T})\ s.t.\ E' \succ^d_{\mathbf{Q},\,\mathsf{d}} E\} = \{E \in \mathbb{E}(\mathbb{T}) \mid \forall E' \in \mathbb{E}(\mathbb{T})\ E \succeq^d_{\mathbf{Q},\,\mathsf{d}} E'\}$（因为 $\succeq^d_{\mathbf{Q},\,\mathsf{d}}$ 是一个全预序）。
  $\mathbb{E}(\mathbb{T}) \mid \forall E' \sigma E \leq \sigma E' \}$. $\forall E \in \mathbb{E}(\mathbb{T})\ \gamma E =$
  因此，$\max(\mathsf{cSuf}(\mathbf{Q}), \mathsf{d}(x_{\downarrow E}, x))$ 且 $\forall E' \notin \mathsf{cSuf}(\mathbf{Q}),\ \gamma(E') = +\infty$，我们有 $\forall E \in \mathsf{cSuf}(\mathbf{Q}),\ \forall E' \notin \mathsf{cSuf}(\mathbf{Q}),\ E \succ^c E'$。由此可得，$\max(\mathbb{E}(\mathbb{T}) \succeq^d_{\mathbf{Q},\mathsf{d}}) \subseteq \mathsf{cSuf}(\mathbf{Q})$。因此，$\max(\mathbb{E}(\mathbb{T}) \succeq^d_{\mathbf{Q},\mathsf{d}}) = \{E \in \mathsf{cSuf}(\mathbf{Q}) \mid \nexists E' \in \mathsf{cSuf}(\mathbf{Q})\ s.t.\ E' \succ^d_{\mathbf{Q},\mathsf{d}} E\} = \mathrm{L}_d(\mathbf{Q})$。

□

定理 13. 考虑查询 $\mathbf{Q} = \langle \mathbb{T},\ \kappa,\ x \rangle$，其中 $\mathbb{T}$ 由布尔特征组成，且 κ 是一个布尔公式。

1. 当$\mathbf{L}$ 为cSuf、sNec或$\mathbf{L_{wf}}$时，问题$\textsc{DecideExp}(\mathbf{L})$ 属于 $O(n)$。
2. 当$\mathbf{L}$为gNec、gSuf、sSuf、$\mathbf{L_c}$、$\mathbf{L}_d$时，问题$\textsc{DecideExp}(\mathbf{L})$ 是 *co-NP* 完全。

对于定义域大小任意的情况，当$\mathbf{L}$ 为cSuf 或$\mathbf{L_{wf}}$时，问题 $\textsc{DecideExp}(\mathbf{L})$ 属于 $O(n)$。对于非布尔域上的查询，当$\mathbf{L}$ 为sNec时，问题$\textsc{DecideExp}(\mathbf{L})$ 是 *co-NP* 完全。

证明.

1 cSuf是所有满足 $\kappa(y) \neq \kappa(x)$的 $y \setminus x$ 集合 .
给定 E，可以在多项式时间内检查是否存在一个实例 y，使得 $E = y \setminus x$ 且 $\kappa(y) \neq \kappa(x)$。事实上，只有在 $x \cap E = \emptyset$ 的情况下，这样的实例 y 才可能存在；此时它必然是 $x_{\downarrow E}$。因此，DECIDEEXP(cSuf) 属于 $O(n)$。

现在我们考虑 sNec .
对于一个 $x, E \in \mathsf{sNec}(\mathbf{Q})$，若对所有满足 $x \setminus y = E$ 的 y，$\kappa(y) \neq \kappa(x)$ 等价于 $E = \overline{\{f_v \mid f_v \in y\}}$。令 $\gamma(E) = \gamma(x_{\downarrow E})$。当所有特征域都是布尔型时，仅对 $y = x_{\downarrow E}$ 成立。因此，对于 $\mathrm{L} = \mathsf{sNec}$，DECIDEEXP(L) 属于 $O(n)$。因为给定 E 后，可以在 $O(n)$ 内检查 $E \subseteq x$ 和 $\kappa(x_{\downarrow E}) \neq \kappa(x)$。现在考虑 $\mathbf{L_{wf}}$。当且仅当 $\kappa(x_{\downarrow E}) \neq \kappa(x)$ 时，$\mathrm{L}_{\mathbf{wf}}(Q)$。DECIDEEXP($\mathbf{L_{wf}}$) 属于 $O(n)$，因为给定 E 后，只需测试 $\kappa(x_{\downarrow E}) \neq \kappa(x)$，这可以在 $O(n)$ 内完成。

2. 对于 gNec、gSuf 和 sSuf 这几种情况，要证明该问题属于 co-NP，只需注意到反例（即满足 $\kappa(y) = \kappa(x)$ 的实例 y）可以在多项式时间内验证。要证明 co-NP 完全性，我们依次考察 gNec、gSuf 和 sSuf。对于每种情况，只需给出一个从 TAUTOLOGY 到该问题的多项式归约；TAUTOLOGY 是判定一个析取范式 $\Psi(f_1, \ldots, f_n)$ 是否为重言式的问题。我们先考察 gNec。定义 $\kappa(f_1, \ldots f_{n+1}) = \Psi(f_1, \ldots, f_n) \wedge f_{n+1}$。令 $x = (0, \ldots, 0)$（于是 $\kappa(x) = 0$），并令 $E = \{(f_{n+1}, 0)\}$.$E \in \mathsf{gNec}(\mathbf{Q})$ 当且仅当对于所有不包含 y 的 $E \subseteq y$，$\kappa(y) \neq \kappa(x)$，也就是说，当且仅当 $\Psi(f_1, \ldots, f_n) \equiv 1$。这是一个从 TAUTOLOGY 到 DECIDEEXP(gNec) 的多项式归约。现在我们考察 gSuf 和 sSuf。我们使用相同的 κ 和 x，但令 $E = \{(f_{n+1}, 1)\}$.$E \in \mathsf{sSuf}(\mathbf{Q})$ 当且仅当 $E \cap x = \emptyset$（这就是此处的情况），且对于所有包含 E 的 y，$\kappa(y) \neq \kappa(x)$。因此，$E \in \mathsf{sSuf}(\mathbf{Q})$ 当且仅当 $\Psi(f_1, \ldots, f_n)$。这就给出了一个从 TAUTOLOGY 到 DECIDEEXP(sSuf) 的多项式归约。gSuf 的证明完全相同。

现在我们考虑 $\mathbf{L_c}$ 和 L_d。$\mathbf{L_c}(\mathbf{Q})$ 和 $\mathrm{L}_d(\mathbf{Q})$ 包含 $\mathsf{cSuf}(\mathbf{Q})$ 中分别按基数或距离 d最小的解释。由于可以在多项式时间内测试是否属于 $\mathsf{cSuf}(\mathbf{Q})$，因此可以在多项式时间内验证反例证书（如果 $E \notin \mathrm{L}$，则证书可以是 E 本身；也可以是另一个优于 E 的 $E' \in \mathrm{L}$）。这表明，当 L 为 $\mathbf{L_c}$ 或 L_d 时，DECIDEEXP(L) 属于 co-NP。为证明 co-NP困难性，考虑 $\kappa(f_1, \ldots, f_{n+1}) = (\Phi(f_1, \ldots, f_n) \wedge f_{n+1}) \vee (f_1 \wedge \ldots \wedge f_{n+1})$，其中 $\Phi(f_1, \ldots, f_n)$ 是合取范式。令 $x = (0, \ldots, 0)$，使得 $\kappa(x) = 0$，并令 $E = \{(f_1, 1) \ldots, (f_{n+1}, 1)\}$。 $E \in \mathsf{cSuf}(\mathbf{Q})$ 且 $E \in \mathrm{L}_{\mathbf{wf}}(\mathbf{Q})$。当且仅当 Φ 可满足时，$\mathsf{cSuf}(\mathbf{Q})$ 中存在另一个解释（若存在，它必然会按基数或距离 d 更受偏好）。因此，我们得到了从不可满足问题到 DECIDEEXP(L) 的多项式归约，其中 L 为 $\mathbf{L_c}$ 和 L_d，从而证明了这两个问题都是 co-NP 困难的。

对于任意定义域大小，当$\mathbf{L}$ 为 cSuf 或$\mathbf{L_{wf}}$ 时，DECIDEEXP($\mathbf{Q}$) 是 $O(n)$ 的证明，与上文布尔情形的证明相同。

现在我们考虑非布尔域上的 DECIDEEXP(sNec)。该问题属于 co-NP，因为反例是一个实例 y，它与 x 恰好在 E 上不同，且满足 $\kappa(y) = \kappa(x)$。为完成 co-NP 完全性的证明，只需给出从 TAUTOLOGY 出发的多项式归约。设 Ψ 是一个包含f_1个变量 $n, \ldots, f_n$ 的析取范式。考虑一个由 n 个特征构成、定义域为 $\{0, 1, 2\}$ 的理论。令 x 为实例 $(2, \ldots, 2)$，并令 $E = x_\diamond$。定义 κ，使得 $\kappa(2, \ldots, 2) = 0$，并且对 $z \in \{0, 1\}^n$ 有 $\kappa(z) = \phi(z)$（其中 κ 在其他实例上取任意值）。那么，$\mathsf{sNec}(\mathbf{Q})$（其中$\mathbf{Q}$是对应于 κ 和 x 的查询）当且仅当 $\Psi \equiv 1$，由此完成了从 TAUTOLOGY 出发的多项式归约。

□

定理 14。 考虑查询 $\mathbf{Q} = \langle \mathbb{T}、\ \kappa、\ x \rangle$，其中 $\mathbb{T}$ 由 n 个布尔特征组成，κ 是一个布尔公式。

1. 当 $\mathbf{L}$ 为 sNec 时，问题 $\textsc{FindExp}(\mathbf{L})$ 为 $O(n)$。*2.* 当 $\mathbf{L}$ 为 gNec、cSuf、sNec、gSuf、$\mathbf{L_c}$、$\mathbf{L}_d$、$\mathbf{L_{wf}}$时，问题 $\textsc{FindExp}(\mathbf{L})$ 为 *NP* 难。在这些情况下，除了 $\mathbf{L_c}$ 和 $\mathbf{L}_d$之外，都至多可通过 n 次对 *SAT* 预言机的调用求解。

对于非布尔域上的查询，当 sSuf 时，$\textsc{FindExp}(\mathbf{L})$ 问题为 *NP* 难。

证明。 1. 我们考虑 sSuf 的情形。令 x 表示 $\{(f, v) \mid (f, v) \in x\}$。如果 $E \in \mathsf{sSuf}(\mathbf{Q})$，则 $E \cap x = \emptyset$，并且对于所有 $y \supseteq E$，$\kappa(y) \neq \kappa(x)$，因此 $\kappa(x \setminus y) \neq \kappa(x)$，所以 $x \in \mathsf{sSuf}(\mathbf{Q})$。因此，要找到 sSuf 中的解释（或证明不存在解释），只需测试 $\kappa(x) \neq \kappa(x)$，其为 $O(n)$。。

2 我们首先考虑 gNec . 设 $\Phi(f_1, \ldots f_n)$ 为一个合取范式 .
定义$\kappa(f_1, \ldots, f_{n+1}) = \Phi(f_1, \ldots, f_n) \vee f_{n+1}$，并令 $x = (1, \ldots, 1)$ 和 $E = \{(f_{n+1}, 1)\}$。 $E \in \mathsf{gNec}(\mathbf{Q})$ 当且仅当对所有 y，$\kappa(y) = \kappa(x)$ 仅当 $E \subseteq y$。换句话说， $E \in \mathsf{gNec}(\mathbf{Q})$ 当且仅当 Φ 不可满足。从 SAT 出发的这一多项式归约证明了 FINDEXP(gNec) 的 NP 困难性。一般而言，寻找某个 $E \in \mathsf{gNec}(\mathbf{Q})$（或判定其不存在），可以通过查找一个单元素的 $E = \{\ell\}$ 来实现，其中 ℓ 是核心文字。由于 κ 是布尔公式，这可以通过 n 次调用 SAT 预言机来实现。

现在我们考虑 cSuf。 $E \in$ cSuf(**Q**) 当且仅当 $E = y \setminus x$ 对某个实例 y 成立，且 $\kappa(y) \neq \kappa(x)$。令 $\Phi(f_1, \ldots, f_n)$ 为一个合取范式。定义 $\kappa(f_1, \ldots, f_{n+1}) = \Phi(f_1, \ldots, f_n) \wedge f_{n+1}$，并令 $x = (0, \ldots, 0)$，使得 $\kappa(x) = 0$。寻找 $E \in$ cSuf(**Q**)（其中**Q** 是相应的查询）相当于为 Φ寻找一个可满足赋值。从 SAT 出发的这一多项式归约证明了 FINDEXP(cSuf) 的 NP 困难性。一般而言，在 cSuf(**Q**) 中寻找一个解释，相当于寻找一个实例 y，使得 $\kappa(y) \neq \kappa(x)$，因此可以通过一次调用 SAT 预言机来实现。

现在考虑 sNec。在布尔域上， $E \in$ sNec(**Q**) 当且仅当 $E \subseteq x$ y $x \setminus y =$，并且存在某个实例，使得 $y \neq \kappa(\Phi)$。令 (f_1, f_n) 为一个合取范式。定义 $\kappa(f_1, \ldots, f_{n+1}) = \Phi(f_1, \ldots, f_n) \wedge f_{n+1}$，并令 $x = (0, \ldots, 0)$，使得 $\kappa(x) = 0$。寻找 $E \in$ sNec(**Q**)（其中 **Q** 是相应的查询）等价于为 Φ 寻找一个可满足赋值。这个从 SAT 到 FINDEXP(sNec) 的多项式归约证明了 FINDEXP(sNec) 的 NP 困难性。

现在考虑 gSuf。 $E \in$ gSuf(**Q**)，如果对于所有包含 E 的 y，都有 $\kappa(y) \neq \kappa(x)$。显然，要找到这样的 E，只需寻找满足 $\kappa(y) \neq \kappa(x)$ 的完全实例 $E = y$。令 $\Phi(f_1, \ldots, f_n)$ 为一个合取范式，并定义 $\kappa(f_1, \ldots, f_{n+1}) = \Phi(f_1, \ldots, f_n) \wedge f_{n+1}$。令 $x = (0, \ldots, 0)$（从而 $\kappa(x) = 0$）。任何满足 $\kappa(y) \neq \kappa(x)$ 的 y，都是由 Φ 给出的 SAT 实例的一个解，这证明了 FINDEXP(gSuf) 的 NP 困难性。
现在我们证明，当**L** 是 gSuf 时，只需调用一次 SAT 预言机，就足以求解 FINDEXP(**L**)。若对于所有包含 E 的 y， $\kappa(y) \neq \kappa(x)$，则称 $E \in$ 为 gSuf(**Q**)。正如我们刚才所指出的，只需寻找一个实例 y，使得 $\kappa(y) \neq \kappa(x)$，而这可以通过调用一次 SAT 预言机实现。

现在我们考虑$\mathbf{L_c}$ 和$\mathbf{L}_d$。由以下证明可直接推出 NP 困难性：当 **L** 为$\mathbf{L_c}$ 或$\mathbf{L}_d$ 时，DECIDEEXP(**L**) 是 co-NP 完全的（定理 13）。
对于$\mathbf{L_{wf}}$的情形，只需考虑 $\kappa(f_1, \ldots, f_{n+1}) = \Phi(f_1, \ldots, f_n) \wedge f_{n+1}$, $x = (0, \ldots, 0)$（因此 $\kappa(x) = 0$）。寻找一个解释 $E \in \mathbf{L_{wf}}$，相当于寻找合取范式 Φ 的一个可满足赋值。因此，DECIDEEXP($\mathbf{L_{wf}}$) 也是 NP 难的。
现在我们证明，当**L** 是$\mathbf{L_{wf}}$时，FINDEXP(**L**)可以通过 SAT 求解。要找到 $E \in \mathbf{L_{wf}}$，只需找到一个实例 y，使得 $\kappa(y) \neq \kappa(x)$，并令 $E = y \setminus x$。这可以通过一次调用 SAT 预言机实现。

对于非布尔域上的查询，我们考虑 sSuf。 $E \in$sSuf(**Q**) 当且仅当 $E \cap x = \emptyset$，且对于所有包含 E 的实例 y，κ x $(y) \neq \kappa(\,)$。因此，如果存在某个 $E \in$sSuf(**Q**)，那么也存在完全实例 $y \in$sSuf(**Q**)。考虑一个由 n 个特征组成、定义域为 $\{0, 1, 2\}$ 的理论。设 $\Phi(f_1, \ldots, f_n)$ 是一个合取范式，令 $x = (2, \ldots, 2)$，并定义 κ，使得 $\kappa(2, \ldots, 2) = 0$，且对于 $z \in \{0, 1\}^n$，$\kappa(z) = \Phi(z)$。那么，寻找一个满足$y \cap x = \emptyset$ 且满足 $\kappa(y) \neq \kappa(x)$ 的实例 y，等价于寻找 Φ 的一个满足实例。由此可知，FINDEXP(sSuf) 是 NP 难问题。 □