arXiv:2304.02781v1 [cs.CC] 5 Apr 2023

决策树的充分理由的不可近似性

Alexander Kozachinskiy[1,2,3]

[1]国家人工智能中心，智利[2]数学与计算工程研究所，智利天主教大学[3]数据基础千年研究所，智利

摘要

在本笔记中，我们确立了计算决策树的 δ-充分原因的最小规模问题的近似困难性。

1 引言

给定人工智能模型的输入，人们可能想要了解模型为何会以这种方式对该输入进行分类。近年来，像这样的局部解释 得到了广泛研究 [8, 7, 6, 10]。其中一种称为充分原因的解释，正是本文的关注重点。

为简单起见，我们将人工智能模型视为布尔函数 $\mathcal{M}$: $\{0,1\}^n \to \{0,1\}$，因此将讨论范围限定为对二元特征向量进行二元分类。 $\mathcal{M}$ 有多种表示方式，例如决策树、神经网络，甚至布尔电路。

对于模型 $\mathcal{M}$: $\{0,1\}^n \to \{0,1\}$ 下的输入 $x = (x_1, \ldots, x_n) \in \{0,1\}^n$，若一组坐标 $S \subseteq \{1, \ldots, n\}$ 满足：对于所有在 S 的所有坐标上都与 x 一致的 $x' \in \{0,1\}^n$，均有 $\mathcal{M}(x') = \mathcal{M}(x)$，则称该组坐标为输入的充分原因 [8, 5]。换言之， x 的充分原因是其一组特征，所有具有这些特征的输入都会被分类为 x。

Barceló 等人 [2] 率先研究了人工智能模型局部解释的 计算复杂性。具体而言，他们提出了最小充分原因 问题。该问题是：给定一个输入 $x \in \{0,1\}^n$ 和一个模型 $\mathcal{M}$: $\{0,1\}^n \to \{0,1\}$，寻找一个关于 x、在 $\mathcal{M}$ 下的充分原因 S，且其规模最小。他们证明，即使 $\mathcal{M}$ 被表示为决策树，该问题也是 NP 困难的（并且其判定版本是 NP 完全的）。

最近，人们注意到 [1, 9] 作为解释的充分理由过于严格，并且在实践中通常太大。我们可能满足于一组更小的特征，这些特征保证与我们的输入 x 在大多数情况下（可能并非总是）具有相同的分类。这导致了对充分原因概念的以下放宽。即，取任意 $\delta \in [0,1]$。一个 δ-充分

原因对于 $x \in \{0,1\}^n$ 在 $\mathcal{M}$下： $\{0,1\}^n \to \{0,1\}$ 是一组坐标 $S \subseteq \{1,2,\ldots,n\}$，满足以下条件：如果我们考虑所有在 S的所有坐标上与 x 重合的 $x' \in \{0,1\}^n$的集合，那么对于这些 x' 中至少分数 δ 的部分，我们有 $\mathcal{M}(x') = \mathcal{M}(x)$。

与"精确"充分理由一样，我们也可以考虑最小 δ-充分理由问题。在该问题中，给定输入 $x \in \{0,1\}^n$ 和模型 $\mathcal{M}$： $\{0,1\}^n \to \{0,1\}$，我们的任务是在 $\mathcal{M}$ 下，为 x找到一个 δ-充分理由 S，且其规模最小。我们假设 δ 是该问题的参数，而不是输入的一部分（因此，对于不同的 $\delta \in [0,1]$，我们得到不同的算法问题）。

Wäldchen 等人 [9] 证明，当 $\mathcal{M}$ 给定为布尔电路时，对于任意 $\delta \in [\frac{1}{2}, 1$, 最小 δ-充分问题的判定版本都是 NP^{PP}完全[1],。另一方面，Arenas 等人 [1] 证明，当 $\mathcal{M}$ 给定为决策树时，对于每个 $\delta \in$ (0,1)，最小 δ-充分理由问题的判定版本都是 NP 完全。

鉴于这些问题难以精确求解，研究其近似困难性是合理的。Wäldchen 等人 [9] 证明了当$\mathcal{M}$ 由布尔电路给定时，最小 δ-充分原因问题具有很强的近似困难性。具体而言，他们证明，区分以下两种情形是 NP 困难的：一种情形是存在一个"非常好"且"非常小"的充分原因；另一种情形是所有"还算不错"的充分原因都包含"几乎所有"坐标。更具体地说，对于每个 $\varepsilon > 0$，他们证明，区分以下两种情形是 NP 困难的：一种情形是存在大小至多为 n^{ε} 的 $(1-\varepsilon)$-充分原因；另一种情形是不存在大小小于 $n - n^{\varepsilon}$ 的 ε-充分原因。这里， n 表示变量的数量。

我们证明，即使 $\mathcal{M}$ 由决策树给定，同样的近似困难性仍然成立。这改进了 Wäldchen 等人的结果，因为每棵决策树都可以在多项式时间内转换为计算相同函数的布尔电路（因此，决策树情形的问题只会比布尔电路情形更容易）。不过，我们采用的复杂性假设比 $P \neq NP$略强。

定理 1. 除非 *SAT* 可在准多项式时间内求解，对于每一个 $\varepsilon > 0$，不存在多项式时间算法，在给定一个在 n 个变量上的决策树 T 及其一个输入 $x \in \{0,1\}^n$ 的情况下，能够区分以下两种情况：

- 存在一个大小至多为 n^{ε} 的 x 在 T 下的 $(1-\varepsilon)$-充分原因
- 不存在大小小于 $n - n^{\varepsilon}$ 的 x 在 T 下的 ε-充分原因。

本文的其余部分组织如下。在第2节中，我们给出预备知识。在第3节中，我们引入一个中间问题，并证明其困难性蕴含定理1。在最后一节中，我们确立中间问题的困难性。

[1]因此，当 $\mathcal{M}$ 由电路给出时，该问题甚至跳出了 NP（当然，除非 $NP = NP^{PP}$）。原因是验证一组坐标是否为 δ-充分原因需要检查某个电路的输入中至少有 δ 的比例是可满足的。

2 预备知识

一个 决策树 T 关于 n 变量 是一个二叉根树，其中

- 每个叶节点都标记为 0 或 1；
- 每个内部节点 v 恰好有两个子节点；其中一个标记为 v 的 0-子节点，另一个标记为 v 的 1-子节点。
- 此外，每个内部节点都标记有一个索引 $i \in \{1,\dots,n\}$。

给定输入 $x=(x_1,\dots,x_n) \in \{0,1\}^n$，$T$ 在 x 上的取值记为 $T(x)$，其计算方式如下。从 T 的根节点出发，我们向下遍历，直到到达某个叶节点。具体而言，如果当前位于内部节点 v，就取其标签 $i \in \{1,\dots,n\}$；若 $x_i = 0$，则向下到达 v 的 0-子节点；若 $x_i = 1$，则向下到达 v 的 1-子节点。到达叶节点后，令 $T(x)$ 等于该叶节点的标签。

除了来自 $\{0,1\}^n$ （“完整”输入）的输入之外，我们还将考虑到 T的部分输入。形式上，部分输入是 $\{0,1,\perp\}^n$ 的一个元素，其中 $\perp$ 是一个表示“未定义”的特殊符号。如果不存在某个 $i \in \{1,\dots,n\}$ 使得 $x_i \neq y_i$和 $x_i, y_i \in \{0,1\}$，我们就称两个部分输入 $x, y \in \{0,1,\perp\}^n$ 是一致的。这意味着，通过将未定义的坐标更改为 0 或1，x 和 y 可以被“补全”为同一个完整输入。

我们将记号 $T(x)$ 扩展到部分输入。具体而言，对于部分输入$y \in \{0,1,\perp\}^n$，我们定义：

$$T(y) = \frac{\left|\{x \in \{0,1\}^n \mid x \text{ and } y \text{ are consistent and } T(x)=1\}\right|}{\left|\{x \in \{0,1\}^n \mid x \text{ and } y \text{ are consistent }\}\right|}.$$

在分母中，我们有 y的 补全 数量，即与 y 一致的完整输入数量。在分子中，我们有满足 $T(x) = 1$的 y 补全数量。因此，$T(y)$ 等于 y 的均匀随机补全使 T等于1的概率。注意，我们可以如下从 y 的补全中均匀随机地采样 $\mathbf{x} \in \{0,1\}^n$：对于所有 $i \in \{1,\dots,n\}$，如果 $y_i = \perp$，则独立地以均匀随机方式选取 $\mathbf{x}_i$，否则将 $\mathbf{x}_i = y_i$ 设为相应值。

我们说 $S \subseteq \{1,2,\dots,n\}$ 是 $x \in \{0,1\}^n$ 在T 下的 δ-充分原因，如果以下成立：

$$\frac{\left|\{z \in \{0,1\}^n \mid z_i = x_i \text{ for every } i \in S \text{ and } T(z) = T(x)\right|}{\left|\{z \in \{0,1\}^n \mid z_i = x_i \text{ for every } i \in S\right|} \geq \delta. \qquad (1)$$

注意到，这一定义也可以用概率术语表述。具体而言，令 $\mathbf{z}$为如下随机变量：对于所有 $i \in \{1,\dots,n\}$，相互独立地进行选择：若 $i \notin S$，则从中均匀随机地选取 $\mathbf{z}_i \in \{0,1\}$；若 $i \in S$，则令$\mathbf{z}_i = x_i$。于是，(1) 可以理解为 $\Pr[T(\mathbf{z}) = T(x)] \geq \delta$。

命题1 (霍夫丁不等式， [4])。 令 X_1， X_2， .. ., X_n 为相互独立的伯努利随机变量。记 $S_n = X_1 + \ldots + X_n$。则对于每个 $\delta > 0$，我们有：

$$\Pr\left[\frac{S_n}{n} \geq \frac{\mathbb{E}S_n}{n} + \delta\right] \leq \exp\{-2\delta^2 n\}.$$

3 中间问题

为证明定理 1，我们引入以下中间问题。给定一个决策树 T，我们希望将 T 中的一些变量固定为1，使 T 等于 1的概率最大化（假设未固定的变量彼此独立且均匀随机地取值）。我们将多少个变量固定为 1 并不重要，关键在于我们不能将任何变量固定为 0。原则上，我们可以将所有变量都固定为1，因此只有当 $T(1, 1, \ldots, 1) = 0$ 时，这个问题才有意义。

定理 2. 除非 *SAT* 可以在准多项式时间内求解，否则对于每个$\kappa > 0$ ，都不存在多项式时间算法，能够给定一个关于 n 个变量的决策树 T，并区分以下两种情况：

- 存在一个部分输入 $y \in \{1, \perp\}^n$ 使得 $T(y) \geq 1 - \kappa$;
- 不存在任何部分输入 $y \in \{1, \perp\}^n$ 使得 $T(y) \geq \kappa$。

该结果的证明将在下一节给出。我们首先由此推导出定理 1。

定理 *2* $\implies$ 定理 *1.* 对于每一个 $\varepsilon > 0$，我们提供一个从定理 2 中带有 $\kappa = \varepsilon/2$ 的问题到定理 1 中带有参数 ε 的问题的多项式时间归约。定义 $m = \lceil (n + \log_2(2/\varepsilon))^{1/\varepsilon} \rceil$。给定一个在 n 个变量上的决策树 T，考虑一个在 $n + m$ 个变量上的决策树 T_1，定义如下：

$$T_1(x_1, \ldots, x_{n+m}) = T(x_1, \ldots, x_n) \vee (x_{n+1} \wedge x_{n+2} \wedge \ldots \wedge x_{n+m}).$$

注意， T_1 可以从 T 在多项式时间内构造（对 T 中每个输出 0 的叶节点附加一棵计算合取 $x_{n+1} \wedge x_{n+2} \wedge \ldots \wedge x_{n+m}$ 的树）。令 e^k 表示一个由 k 个 1 组成的布尔向量。剩下的是建立两个断言。

- 如果存在一个部分输入 $y \in \{1, \perp\}^n$ 使得 $T(y) \geq 1 - \kappa$，那么存在一个大小至多为 $(n + m)^\varepsilon$ 的 $(1 - \varepsilon)$-充分原因，用于 T_1 下的 e^{n+m}。

- 如果不存在部分输入 $y \in \{1, \perp\}^n$ 使得 $T(y) \geq \kappa$，那么不存在 ε-充分原因，用于在 T_1 下为 e^{n+m} ，其大小小于 $(n + m) - (n + m)^\varepsilon$。

让我们从第一个断言开始。首先，注意到 $T(e^{n+m}) = 1$。定义 $S = \{i \in \{1, 2, \ldots, n\} \mid y_i = 1\}$。$S$的大小至多为 $n = (n^{1/\varepsilon})^{\varepsilon} \leq m^{\varepsilon} \leq (n+m)^{\varepsilon}$。还需证明，$S$是在 T_1 下关于 e^{n+m} 的 $(1-\varepsilon)$-充分原因。换言之，我们必须证明 $\Pr[T_1(\mathbf{z}) = 1] \geq 1 - \varepsilon$，其中 $\mathbf{z}$ 是如下随机变量：对于所有 $i \in \{1, \ldots, n+m\}$，若 $i \notin S$，则独立地以均匀随机方式选取 $\mathbf{z}_i$，否则将 $\mathbf{z}_i = 1$ 设为相应值。注意，$\mathbf{z}_1 \ldots \mathbf{z}_n$ 是 y 的均匀随机补全。因此，$T(\mathbf{z}_1 \ldots \mathbf{z}_n) = 1$ 的概率为 $T(y) \geq 1 - \kappa \geq 1 - \varepsilon$。还需注意，$T(\mathbf{z}_1 \ldots \mathbf{z}_n) = 1 \implies T_1(\mathbf{z}) = 1$。

现在，让我们证明第二个结论。假设 $S \subseteq \{1, \ldots, n+m\}$ 是 ε在 T_1 下的 e^{n+m} -充分原因。我们的目标是证明 S的大小至少为 $(n+m) - (n+m)^{\varepsilon}$。同样地，令$\mathbf{z}$为如下随机变量：对于所有 $i \in \{1, \ldots, n+m\}$，如果 $i \notin S$，则独立地以均匀随机方式选取 $\mathbf{z}_i$，否则将其设为 $\mathbf{z}_i = e_i^{n+m} = 1$。$S$ 是 ε在 T_1 下的 e^{n+m} -充分原因这一事实意味着$\Pr[T_1(\mathbf{z}) = 1] \geq \varepsilon$。根据 T_1 的定义以及并集界，我们有：

$$\varepsilon \leq \Pr[T_1(\mathbf{z}) = 1] \leq \Pr[T(\mathbf{z}_1 \ldots \mathbf{z}_n) = 1] + \Pr[\mathbf{z}_{n+1} \wedge \ldots \wedge \mathbf{z}_{n+m} = 1] \qquad (2)$$

注意，$\mathbf{z}_1 \ldots \mathbf{z}_n$ 是 $y \in \{1, \perp\}^n$ 的均匀随机补全，其定义为

$$y_i = \begin{cases} 1 & i \in S, \\ \perp & i \notin S, \end{cases} \qquad i \in \{1, \ldots, n\}.$$

因此，$\Pr[T(\mathbf{z}_1 \ldots \mathbf{z}_n) = 1] = T(y) < \kappa = \varepsilon/2$。于是，由 (2) 可得 $\Pr[\mathbf{z}_{n+1} \wedge \ldots \wedge \mathbf{z}_{n+m} = 1] > \varepsilon/2$。由于 $\mathbf{z}$ 的各坐标相互独立，我们有 $\Pr[\mathbf{z}_{n+1} \wedge \ldots \wedge \mathbf{z}_{n+m} = 1] = \Pr[\mathbf{z}_{n+1} = 1] \cdot \ldots \cdot \Pr[\mathbf{z}_{n+m} = 1]$。根据定义，$\Pr[\mathbf{z}_i = 1] = 1/2 \quad i \notin S \qquad [_{\ i} = 1] = 1 \quad i \in S \qquad \varepsilon/2 < \quad [_{\ n+1} \wedge$ if 且 Pr z if 。$\ldots \wedge _{\ n+m} = 1] = 2^{-|\{n+1, \ldots, n+m\} \setminus S|} \qquad |\{n+1, \ldots, n+m\} \setminus S| <$ 这意味着 Pr z z 。由此可得 $\log_2(2/\varepsilon)$。特别地，我们得到 $|S| \geq m - \log_2(2/\varepsilon)$。还需证明 $m - \log_2(2/\varepsilon) \geq (n+m) - (n+m)^{\varepsilon}$，或者等价地，$(n+m)^{\varepsilon} \geq n + \log_2(2/\varepsilon)$。事实上，$(n+m)^{\varepsilon} \geq m^{\varepsilon} = \lceil (n + \log_2(2/\varepsilon))^{1/\varepsilon} \rceil^{\varepsilon} \geq n + \log_2(2/\varepsilon)$。 □

Wäldchen 等人 [9] 隐式地在 T 为布尔电路时建立了定理 2。随后，与上述相同的归约为他们关于布尔电路的难以近似性结果提供了证明。

下面是他们论证的概述。该归约来自 CNF 的可满足性问题。给定一个 CNF C，我们将其中的每个变量替换为由足够长的全新变量构成的合取式。令所得电路为 T。假设 T 是关于 N 个变量的。如果 C 不可满足，则对于每个 $y \in \{1, \perp\}^N$，$T(y) = 0$。现在假设 C 是可满足的。我们取 C 的任意满足赋值 α，并将其转换为 $y \in \{1, \perp\}^N$，使得 $T(y)$ 非常接近 1。具体而言，如果 C 中的某个变量在 α 中被赋值为 1，我们就将相应合取式中的所有变量都固定为 1。否则，我们将该合取式中的所有变量都保留为未定义。以高概率，每个合取式的值都将等于 α 中对应变量的值。因此，以高概率，T 在 y 的随机补全上的值将等于 $C(\alpha) = 1$。

我们现在针对决策树提出一个论证。主要挑战在于，决策树的可满足性问题可以在多项式时间内求解。因此，我们必须基于不同的思路来确立这一问题的困难性。

4 定理2的证明

我们从 1-in-k 精确命中集问题（1-in-EkHS）归约而来。这是 SAT 问题的一个变体，其中每个子句恰好包含 k 个变量（不含否定），且当且仅当其中恰有一个变量被设为 1 时，该子句才满足。我们使用 Guruswami 和 Trevisan [3, 引理 13] 的以下结果：

定理3。对于每个 $\delta > 0$，存在 k，使得区分 *1-in-EkHS*的可满足实例与 *1-in-EkHS* 的实例是*NP* 困难的；对于后者，不可能满足超过 $(1/e + \delta)$ 的子句比例。

我们固定任意 k，使得区分 1-in-EkHS的可满足实例与1-in-EkHS 的实例在NP困难性上成立；对于后者，不可能满足超过子句的 fraction $1/2$ 。现在，固定 $\kappa > 0$。为证明定理 2，我们取任意一个具有 n 个变量和 m 个子句的 1-in-EkHS 实例 φ，并在准多项式时间内构造一个关于 N 个变量的决策树 T，使得

- 若 φ 可满足，则存在 $y \in \{1, \perp\}^N$ 使得 $T(y) \geq 1 - \kappa$；
- 如果不可能满足 φ的子句中超过 $1/2$的比例，那么不存在 $y \in \{1, \perp\}^N$ 使得 $T(y) \geq \kappa$。

令 l 为满足 $m \leq 2^{2l}$ 的最小整数。注意，$l = \Theta(\log m)$。证明中技术性最强的部分是在多项式时间内构造一个 $O(\log m)$-深度决策树 L，它基于 $n + 2l + 2$ 个变量，使得对于某个绝对常数 $c > 0$，我们有：

- 若 φ 是可满足的，则存在 $y \in \{1, \perp\}^{n+2l+2}$ 使得 $L(y) \geq 7/8$；
- 如果无法满足超过 $1/2$比例的 φ子句，那么对于每个 $y \in \{1, \perp\}^{n+2l+2}$，都有 $L(y) \leq 7/8 - \delta$，其中$\delta = \frac{c}{\sqrt{\ln(m)}}$。

我们首先描述如何将“间隙”从($7/8 - \delta$ vs. $7/8$)增大到(κ vs. $1 - \kappa$)。 L 的构造将在之后给出。

完成证明，模去 L的构造。我们定义一个决策树 T 为

$$K = \frac{2\ln(2/\kappa)}{\delta^2} = O(\log m)$$

“ L”的独立副本。更具体地说，我们令 T作用于 $N = K(n + 2l + 2)$个变量上。首先，T在前 $n + 2l + 2$ 个变量上运行 L，然后在第二

$n+2l+2$变量，依此类推（总计运行 K 次）。最后，当且仅当至少有 $7/8-\delta/2$ 的运行中 L 的输出为 1 时，它输出 1。

首先假设 φ 是可满足的。那么，根据 L 的定义，存在 $y \in \{1,\perp\}^{n+2l+1}$，使得 $L(y) \geq 7/8$。重复 K 次即可得到 T 的一个部分输入$Y \in \{1,\perp\}^N$。我们声称，$T(Y) \geq 1-\kappa$。事实上，可以将 Y 的一次随机补全生成为 y 的一次随机补全的 K 个独立样本。树 T 输出 1，当且仅当 L 在至少 $7/8-\delta/2$ 的样本中输出 1。另一方面，L 在某个样本上输出 1 的概率为 $L(y) \geq 7/8$。因此，L 输出 1 的样本平均比例至少为 $7/8$。根据霍夫丁不等式（命题 1），T 输出 1 的概率至少为 $1-\exp\{-2(\delta/2)^2 \cdot K\} = 1-\kappa/2 > 1-\kappa$。

现在，假设不可能满足超过 $1/2$ 个 φ 的子句的比例。于是，对于所有 $y \in \{1,\perp\}^{n+2l+2}$，$L(y) \leq 7/8-\delta$。我们声称，对于每个 $Y \in \{1,\perp\}^N$，$T(Y) < \kappa$。令 y_1 为将 Y限制到前 $n+2l+2$ 个变量上的限制，令 y_2 为将 Y 限制到接下来 $n+2l+2$ 个变量上的限制，依此类推。对 Y 生成随机补全，等同于独立地对 y_1、y_2、.....、y_K 采样随机补全。对于每个 $i \in \{1,\ldots,K\}$，L 在 y_i 的随机补全上输出 1 的概率为 $L(y_i) \leq 7/8-\delta$。因此，满足以下条件的 $i \in \{1,\ldots,K\}$ 的平均比例至多为 $7/8-\delta$：L 在 y_i 的随机补全上输出 1。反过来，$T(Y)$是满足以下条件的 $i \in \{1,\ldots,K\}$ 的比例至少为 $7/8$ 的概率：L在 y_i 的随机补全上输出 1。因此，根据霍夫丁不等式，$T(Y)$ 被 $\exp\{-2(\delta/2)^2 \cdot K\} = \kappa/2 < \kappa$ 所界定。

T的深度是(L的深度) $\times K = O(\log^2 m)$。因此，T的大小是准多项式。构造 T 从 L 的时间也是准多项式，因为它是 T大小的多项式。

构造 L。令 $x_1,\ldots,x_n$表示 φ的变量。L 的变量将记为 $x_1,\ldots,x_n,y_1,\ldots,y_{2l+1},z$。

首先，L询问 y_1、.....、y_{2l+1}的取值。现在，如果一个二进制词中1的数量多于0，我们就称其为“胖”，否则称其为“瘦”。如果 $y_1 \ldots y_{2l+1}$是瘦的，L输出1。为了描述 $y_1 \ldots y_{2k+1}$为胖时 L 的行为，我们首先注意到，在 $\{0,1\}^{2k+1}$ 中，恰好一半的词是胖的，恰好一半的词是瘦的。特别地，胖词的数量为 2^{2l}。我们固定一个从 $\{0,1\}^{2l+1}$ 中的胖词集合到 φ的子句集合的满射映射。这样的映射存在，这是因为根据我们对 l的选择，$m \leq 2^{2l}$。

因此，当 $y_1 \ldots y_{2l+1}$胖时，L取子句 C，将其赋给 $y_1 \ldots y_{2l+1}$，并运行决策树 L_C，定义如下。如果 C涉及变量 x_{i_1}、...、x_{i_k}，那么 L_C涉及 x_{i_1}、..、x_{i_k}和 z。树 L_C询问其所有变量的值。当且仅当 x_{i_1}、...、x_{i_k}中恰好有一个 0，或（$x_{i_1} = \ldots = x_{i_k} = 1$和 $z = 0$）时，它输出 1。

L的深度为 $2l+1+k+1 = O(l) = O(\log m)$，且可在多项式时间内由 φ 构造。

引理1。考虑任意部分输入 p到 L_C，其中没有任何坐标等于 *0*（也就是说，只能有取值固定为 *1* 的变量和未赋值变量）。

如果 z 在 p 中是未赋值的，并且 $x_{i_1}, \ldots, x_{i_k}$ 中恰好还有一个未赋值变量，则称 p 为好。否则，称 p 为坏。

那么，如果 p 是好的，我们有 $L_C(p) = 3/4$，如果 p 是坏的，我们有 $L_C(p) \leq 5/8$。

证明。首先假设 p 是好的。令 x_{i_j} 为 p 中未定义的 x_{i_1}、……、x_{i_k} 之一。考虑 p 的一个随机补全。如果在此补全中 x_{i_j} 为 0，那么 x_{i_1}、……、x_{i_k} 中恰好有一个为 0，因此 L_C 输出 1。如果 x_{i_j} 为 1，那么 L_C 以概率 $1/2$ 输出 1，具体取决于 $z = 0$。总之，我们得到 $L_C(p) = (1/2) \cdot 1 + (1/2) \cdot (1/2) = 3/4$。

现在，考虑 p 很糟糕的情况。首先，假设在 p中的 x_{i_1}、……、x_{i_k} 之间有 $t \geq 2$个未赋值变量。考虑p的一个随机补全。根据定义，L_C 仅在以下两种情况下对其输出1：（a）在 x_{i_1}、……、x_{i_k} 之间恰好有1个未赋值变量在我们的随机补全中等于0；（b）x_{i_1}、……、x_{i_k} 之间的所有未赋值变量都等于1，并且 $z = 0$ 在我们的随机补全中。情况（a）的概率为 $t2^{-t}$。如果 z 是未赋值的，则情况（b）的概率为 2^{-t-1}，否则为0。总的来说，我们得到$L_C(p) \leq t2^{-t} + 2^{-t-1}$。该表达式随 t递减，并且当 $t = 2$ 时等于 $5/8$。

现在，假设 x_{i_1}，...，x_{i_k} 中的未赋值变量数至多为 1。若该数量为 1，则在 p 中必须将 z 固定为 1（否则，p 就是好的）。此时，只有当 x_{i_1}，...，x_{i_k} 中唯一的未赋值变量为 0 时，L_C 才能输出 1；在对 p 进行随机补全时，这种情况发生的概率为 $1/2$。也就是说，在这种情况下，$L_C(p) = 1/2$。现在，如果在 p 中，所有变量 x_{i_1}，...，x_{i_k} 都被固定为 1，那么只有当 $z = 0$ 时，L_C 才会输出 1。如果 z 未赋值，其概率为 $1/2$；如果 z 被固定为 1，则概率为 0。因此，在这种情况下，$L_C(p) \leq 1/2$。□

首先假设 φ 是可满足的。固定一个关于 φ 的满足赋值 α。我们构造一个关于 L 的部分输入 $p \in \{1, \perp\}^{n+2l+2}$，使得 $L(p) = 7/8$。我们在 x_1，...，x_n 上如下定义 p：

$$x_i = 0 \text{ in } \alpha \implies x_i = 1 \text{ in } p \tag{3}$$

$$x_i = 1 \text{ in } \alpha \implies x_i = \perp \text{ in } p \tag{4}$$

变量 y_1、……、y_{2l+1}、z 在 p 中未定义。由于 α 是 φ 的满足赋值，因此对于每个子句 C，C 中恰好有一个变量在 p 中未定义。这意味着，对于 C 的每个子句，将 p 限制在 L_C 的变量上所得的限制具有引理1所述的良好性质。因此，在“$y_1 \ldots y_{2l+1}$ 是胖的”这一事件的条件下，在对 p 进行随机补全时，L 以概率 $3/4$ 输出 1。现在，变量 y_1、……、y_{2l+1} 在 p 中未定义，这意味着在我们的补全中，它们都是相互独立且均匀地随机抽样的。事件“$y_1 \ldots y_{2l+1}$ 是胖的”的概率为 $1/2$，这对 $L(p)$ 的贡献为 $(1/2) \cdot (3/4) = 3/8$。反过来，在事件“$y_1 \ldots y_{2l+1}$ 是瘦的”（该事件发生的概率同样为 $1/2$）的条件下，我们的树总是输出 1。总的来说，我们得到 $L(p) = 3/8 + 1/2 = 7/8$。

现在我们证明，如果无法满足 φ中超过一半的子句，那么对于每个 $p \in \{1, \perp\}^{n+2l+2}$ 我们有 $L(p) \leq 7/8 - \delta$，其中 $\delta = c/\sqrt{\ln(m)}$

且 $c > 0$ 是某个绝对常数。同样地，我们通过对 p 进行随机补全，并证明 L 在其上输出 1 的概率至多为 $7/8 - \delta$，来实现这一点。

首先假设 $y_1, ..., y_{2l+1}$ 在 p 中均为未定义。我们通过反转 (3–4) 将 p 转换为对 φ 的赋值 α：

$$x_i = 1 \text{ in } p \implies x_i = 0 \text{ in } \alpha$$
$$x_i = \perp \text{ in } p \implies x_i = 1 \text{ in } \alpha$$

请注意，如果 α 不满足子句 C，那么 p 在 L_C 的变量上的限制，从引理1的意义上说是坏的。现在，未满足子句的数量至少为 $m/2$。这些子句分别归属于某个胖 $y_1 \dots y_{2l+1}$。也就是说，对于至少 $m/2$ 个胖 $y_1 \dots y_{2k+1}$，在固定 $y_1 \dots y_{2k+1}$的条件下，L 输出1的概率至多为 $5/8$。对于任何其他胖 $y_1 \dots y_{2k+1}$，这个条件概率至多为 $3/4$。总的来说，胖$y_1 \dots y_{2k+1}$ 对 $L(p)$的贡献不超过

$$\frac{m/2}{2^{2l+1}} \cdot (5/8) + \frac{2^{2l} - m/2}{2^{2l+1}} \cdot (3/4),$$

以及瘦 $y_1 \dots y_{2k+1}$ 对 $L(p)$ 的贡献，和之前一样，是 $1/2$。总体而言，我们得到

$$L(p) \leq \frac{m/2}{2^{2l+1}} \cdot (5/8) + \frac{2^{2l} - m/2}{2^{2l+1}} \cdot (3/4) + 1/2 = 7/8 - \frac{m}{2^{2l+5}}.$$

还需回顾，l 是选取的满足$m \leq 2^{2l}$的最小整数。这意味着 $m \geq 2^{2(l-1)}$。由此可得 $L(p) \leq 7/8 - 1/128 < 7/8 - \delta$。

还需考虑以下情况：在 p 中，y_1、……、y_{2l+1}中至少有一个变量被固定为 1。与之前一样，在“y_1、……、y_{2k+1}胖”这一事件的条件下，L 输出1 的概率至多为 $3/4$。而在“y_1、……、y_{2k+1}瘦”这一事件的条件下，该概率为 1。然而现在，由于 y_1、……、y_{2k+1}中至少有一个变量被固定为 1，取瘦的概率略小于 $1/2$，这使得 $L(p)$ 略小于 $7/8$。更具体地说，

$$\begin{aligned} L(p) &\leq \Pr[y_1 \dots y_{2l+1} \text{ is thin}] + (3/4) \Pr[y_1 \dots y_{2l+1} \text{ is fat}] \\ &= 1 - (1/4) \Pr[y_1 \dots y_{2l+1} \text{ is fat}], \end{aligned}$$

其中 y_1、……、y_{2l+1} 从 p 的随机补全中选取。也就是说，现在我们的目标是在 y_1、……、y_{2l+1} 中至少有一个被固定为 1、其余变量独立且均匀随机选取的前提下，给出 $\Pr[y_1 \dots y_{2l+1}$为胖$]$, 的下界。我们固定为 1 的变量越多，$\Pr[y_1 \dots y_{2l+1}$为胖$]$ 就越大。因此，不失一般性，我们可以假设恰好有一个变量被固定为 1，例如 y_{2l+1}。当且仅当 $y_1 \dots y_{2l+1}$ 中至少有 l

个 1 出现在 y_1、.....、y_{2l} 中时，为胖。因此，我们得到

$$\begin{aligned}\Pr[y_1 \dots y_{2l+1} \text{ is fat}] &\geq \frac{\binom{2l}{2l} + \dots + \binom{2l}{l+1} + \binom{2l}{l}}{2^{2l}} \\ &= \frac{\frac{\binom{2l}{2l}+\binom{2l}{0}}{2} + \dots + \frac{\binom{2l}{l+1}+\binom{2l}{l-1}}{2} + \frac{\binom{2l}{l}+\binom{2l}{l}}{2}}{2^{2l}} \\ &= \frac{\binom{2l}{2l} + \dots + \binom{2l}{0}}{2 \cdot 2^{2l}} + \frac{\binom{2l}{l}}{2 \cdot 2^{2l}} = \frac{1}{2} + \Omega(1/\sqrt{l}) \\ &= \frac{1}{2} + \Omega(1/\sqrt{\ln m}).\end{aligned}$$

由此，我们得到 $L(p) \leq 1 - (1/4) \cdot (1/2 + \Omega(1/\sqrt{\ln m})) = 7/8 - \Omega(1/\sqrt{\ln m})$，符合要求。

参考文献

[1] Arenas, M.、Barceló, P.、Romero Orth, M. 和 Subercaseaux,B. 关于计算决策树的概率解释。 神经信息处理系统进展 *35* （2022），28695–28707。[2] Barceló, P.、Monet, M.、Pérez, J. 和 Subercaseaux, B.通过计算复杂性的视角理解模型可解释性。 神经信息处理系统进展 *33* （2020），15487–15498。[3] Guruswami, V. 和 Trevisan, L.唯一选择的复杂性：近似 1-in-k 可满足性问题。载于第八届近似、随机化与组合优化问题国际研讨会论文集，以及第九届随机化与计算国际会议：算法与技术论文集 （2005），第 99–110 页。[4] Hoeffding, W. 有界随机变量之和的概率不等式。 美国统计学会期刊 *58*，301（1963），13–30。[5] Ignatiev, A. 和 Marques-Silva, J. 基于 SAT 的严格解释：决策列表。载于 可满足性测试理论与应用—*SAT2021*：第 *24* 届国际会议，西班牙巴塞罗那，*2021* 年 *7* 月 *5—9*日，论文集 *24* （2021），施普普林格，第 251–269 页。[6] Lundberg, S. M. 和 Lee, S.-I. 解释模型预测的统一方法。 神经信息处理系统进展 *30*（2017）。[7] Ribeiro, M. T.、Singh, S. 和 Guestrin, C.Anchors：高精度的模型无关解释。载于 *AAAI* 人工智能会议 （2018），第 32 卷。

[8] Shih, A., Choi, A. 和 Darwiche, A. 一种解释贝叶斯网络分类器的符号方法。载于第 *27* 届国际人工智能联合会议论文集（2018），第 5103–5111 页。

[9]Wäldchen, S., Macdonald, J., Hauch, S. 和 Kutyniok, G. 理解二元分类器决策的计算复杂性。《人工智能研究杂志》第 *70* 卷（2021），351–387。

[10] Yan, T. 和 Procaccia, A. D. 如果你喜欢 Shapley，你一定会爱上核心。在 *AAAI* 人工智能会议论文集（2021），第 35 卷，第 5751–5759 页。