arXiv:2305.01063v2 [cs.AI] 4 May 2023

专长树　　解决集体决策中的知识局限

Axel Abels [1 2] **Tom Lenaerts** [1 2 3] **Vito Trianni** [4] **Ann Now´e** [2]

摘要

为决策者提供建议的专家，其展现出的专长可能随问题实例而变化。在实践中，这可能导致针对少数群体案例的次优决策或歧视性决策。在本研究中，我们将知识深度和广度的此类变化建模为问题空间的划分，将其划分为专长各异的区域。我们提出了新的算法，明确考虑并适应问题实例与专家知识之间的关系。首先，我们提出了一种基于近邻查询的朴素方法，并指出其缺陷。为解决这些缺陷，我们随后引入一种新颖的算法——专长树。该算法构建决策树，使学习器能够选择适当的模型。我们给出理论分析，并通过实证验证了这一新方法在一系列现有方法表现不足的问题上的性能提升。

1. 引言

以下医学诊断示例说明了本文所研究的问题设定：一个在线平台反复向一组医学专家展示具有不同症状的患者。必须根据这些症状、患者的病历以及专家的知识和经验，选择最合适的治疗方案。随后对患者实施该治疗方案，并根据康复情况、副作用和成本等指标评估其有效性。在这种设定下，一个合理的目标是最大限度地提高所实施治疗方案的有效性，具体取决于这组医学专家的建议。理想情况下，我们希望医学专家根据每种治疗方案的客观特征（例如成功率、副作用和成本）就最佳治疗方案达成一致。然而，专家的意见可能同时受到客观因素和主观观点的影响。例如，临床医生可能会根据以往治疗类似患者的经验，对某种治疗方案对某位患者的疗效形成看法，而另一位临床医生的看法可能与之不同。这种分歧可能源于显然超出专家控制范围的因素，但也可能是认知偏差所致。例如，医学专家可能会受到确认偏误的影响，从而夸大某种治疗方案的疗效（Trope etal., 1997）或首因效应（Bondet al., 2007）。因此，专家会根据以往处理类似问题的经验提供个人建议。当我们能够接触到一组意见可能不一致的专家时，就需要决定采纳哪项建议。具体而言，我们考虑决策过程经过多次迭代的情形（针对不同的问题实例，例如具有不同病历的不同患者），以便根据专家专长进行调整。由于反馈不完整，确定如何采纳专家建议面临更大挑战：我们只能观察自己所做决策的结果，例如所选的治疗方案。

面对不确定性，利用专家建议解决复杂决策问题的做法已得到广泛研究，这类问题称为多臂老虎机（Auer 等人，2002a；Foster 和 Rakhlin，2020）。然而，以往在此情境下利用专家建议的方法，并未根据各个问题实例的特征进行调整。与此不同，我们在此考虑局部专长，具体而言，专家的水平可能在每次迭代中随问题实例（即观测到的一组特征）而变化。对于人类专家来说，这种专长水平的差异很常见，因为他们通常专精于更大问题领域中的特定部分。同样，人工专家也可能因训练不足，或数据集不完整或存在偏差而出现性能波动。具体来说，（有意或无意的）偏差可能导致判断错误，例如对性别或族裔等敏感特征赋予不恰当的权重（Pohl，2017年）。无论是人类专家还是人工专家，都可能因有偏设计或数据偏差而产生此类偏差（Gianfrancesco 等人，2018）。能够检

[1]布鲁塞尔自由大学机器学习组，布鲁塞尔，比利时 [2]布鲁塞尔自由大学人工智能实验室 [3]加州大学伯克利分校人类兼容人工智能中心，伯克利，美国 [4]国家研究委员会认知科学与技术研究所，罗马，意大利。通信作者：Axel Abels< axel.abels@ulb.be>。

被 40th 机器学习国际会议接收，檀香山，夏威夷，美国。PMLR 202，2023年。版权所有 2023年，归作者所有。

测并纠正此类偏差的方法，不仅对性能至关重要，对公平性也同样重要。具体而言，我们探讨了在考虑专长随一组特征而变化（例如因专业化或偏差所致）的情况下，性能能够得到怎样的提升。因此，尽管可能不存在一种能够全局最优地利用专家建议的方式，但划分问题空间，并针对这些划分建立局部模型，或许能提供有效的分段模型。本文认为，为集体决策设计的算法必须考虑局部专长，才能最大限度地提升性能。尤其是，专家建议多臂老虎机算法应学习根据问题实例进行调整的策略。

为满足这一需求，我们提出两种新颖的算法：一种基于最近邻方法，另一种基于模型树方法。为凸显其与模型树的区别，我们将后一种方法称为**专业知识树**。这些方法日益增加的复杂度，是应对局部专长挑战所必需的。随后，我们给出局部情境下算法性能的一些界，并从理论上分析不恰当划分问题空间所造成的损失。最后，我们通过实证评估这些算法的性能和执行时间，结果表明，专业知识树方法取得了最佳效果，且算法复杂度的增幅尚可接受。

2. 局部专家建议多臂老虎机

多臂老虎机将这类问题形式化：学习器在若干轮次中反复从 K 个臂中选择一个，目标是在只能观测到所选臂结果的情况下，优化其选择的结果（Thompson，1933；Auer 等，2002a）。在上下文情境中（Chu 等，2011；Valko 等，2013），结果由函数 f：$[K] \times \mathrm{R}^d \to [0,1]$表征，该函数将一个臂$k$ 和决策在时间 t 的上下文 $\vec{x}_t$ （即表征当前实例的一组特征）映射为期望奖励。在时间 t，在上下文 $\vec{x}_t$ 下选择臂k_t所观测到的奖励为 $r_t = f(k_t, \vec{x}_t) + \epsilon$，其中 ϵ 是均值为 0 的噪声项。例如，就 QALY（Whitehead 和 Ali，2010）等相关指标而言，结果（奖励）最佳的治疗方案（臂）会随患者病历（上下文）而变化。当问题的复杂程度导致学习成本难以接受时，学习器可以利用专家的建议来做出决策，这一问题由专家建议多臂老虎机形式化（Auer 等，2002a；Foster 和 Rakhlin，2020）。具体来说，当从头近似 f 不可行时，学习器可以观测一组 N 专家提供的建议，其形式为对每个臂的期望奖励的 N 估计值。我们用 $\vec{\xi}^n : \mathrm{R}^d \to \mathrm{R}^K$ 表示专家 n的...

K 估计值的向量函数，每个臂对应一个。

为了最大化所选各臂带来的奖励，学习器必须根据这些建议向量优化决策。因此，带有专家建议的多臂老虎机算法会通过维护由建议矩阵 $\boldsymbol{\xi}(\vec{x}_t) := \{\vec{\xi}^1(\vec{x}_t), \ldots, \vec{\xi}^N(\vec{x}_t)\}$ 参数化的策略 π_t，来学习如何利用建议矩阵采取行动。为便于阅读，在不存在歧义时，我们省略 ξ 的参数；例如，我们用 $\boldsymbol{\xi}_t$ 简写由时间 t 的上下文诱导的 $N \times K$ 建议矩阵。类似地，$\xi_{k,t}$ 表示该矩阵的第 k 列，即针对臂 k 的建议。在每一轮 t 中，策略 π_t 会选择一个臂 k_{π_t}，并观测到奖励 r_{π_t}。学习器的目标是根据观测到的奖励更新策略，以最大化收集到的奖励的平均值 $\frac{1}{T}\sum_{t=1}^{T} r_t$。给定一个最优策略 π^*，性能也可以表示为相对于该策略的期望差距，即遗憾：$R(T) = \mathrm{E}[\frac{1}{T}\sum_{t=1}^{T} r_{\pi^*} - r_{\pi_t}]$，其中期望是针对所有可能的上下文及其相应结果计算的。

2.1. 局部专长

我们通过明确考虑这样一些情况来推广这一问题：专家建议的质量会随着其所面对的上下文而发生变化。对于通常具有特定专业领域的人类专家而言，这种情况自然存在。当某个实例超出专家的专业领域时，就应当将这一点纳入考量。同样，人工专家也可能仅适用于问题空间的某个子集。无论是否出于有意识的原因，偏差都可能导致模型受到某些敏感特征（如性别或族裔）的影响，从而产生不准确的结果。识别专家质量对一组特征的依赖关系，有望显著提高集体决策的质量。我们假设，可以通过划分上下文空间，并在这些划分上学习局部模型，从而获得一个有效的模型。

我们通过引入**专业知识上下文**来形式化这种依赖关系，它被定义为专家观测到的完整上下文的一个子集 $\vec{x}_t$。例如，患者的性别可能会影响治疗方案的有效性，也可能（在无意识中）影响专长的质量。虽然在决定采纳哪种专长时，合理的做法或许是纳入完整上下文 $\vec{x}_t$，但这种子集表述考虑到了某些特征无法被有意义地捕捉的情况。例如，人类专家可以轻易地利用问题的文本描述，但对专家建议多臂老虎机算法而言，考虑这些描述的计算成本可能过高。我们将维度为 $g \leq d$的子集记作 $\vec{z}_t \in \mathrm{R}^g$。因此，在局部专长设定中，采纳专家建议的最优方式取决于这一特征子集。例如，如果患者的症状与心脏有关，就应优先考虑具有相应专长的医学专家的建议，

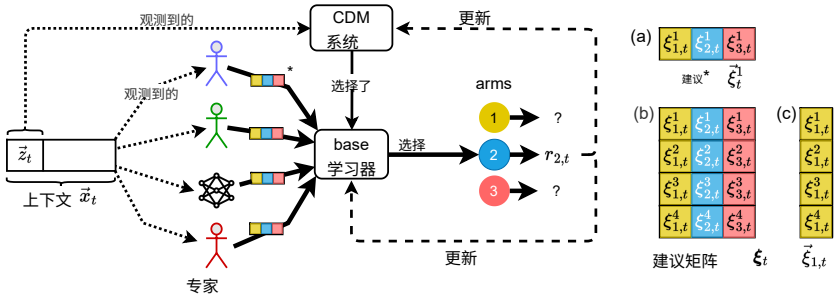

图 1：具有局部专长的多臂老虎机示意图，其中包含 $N = 4$ 位专家和 $K = 3$ 条臂。集体决策系统观察专业知识上下文 $\vec{z}_t$，并选择适当的局部学习器。每位专家在时间 t 观察完整上下文 $\vec{x}_t$，然后向学习器提供一个建议向量（例如，(a) $\vec{\xi}_t^1$ 是第一位专家的建议）。请注意，如图所示，专家既可以是人类，也可以是人工专家。建议矩阵 (b) $\boldsymbol{\xi}_t$ 由这些建议向量拼接而成，而 (c) $\vec{\xi}_{1,t}$ 是针对臂 1 的建议。所选臂（$k = 2$）提供奖励 $r_{2,t}$，该奖励从均值为 $f(2, \vec{x}_t)$ 的分布中采样得到。随后，利用该奖励更新学习器如何依据建议采取行动，以及集体决策系统如何选择合适的学习器。

其专科方向与之匹配。实际而言，$\vec{z}_t$ 编码了患者的类别，决策者应根据 $\vec{z}_t$所编码的类别，优化对专家建议的利用。我们尤其关注这样一种设定：专业知识上下文$\vec{z}$与专家之间的关系事先未知，因此必须通过决策反馈来学习。

2.2. 化简为上下文老虎机

作为基线，我们考虑在每个时间步选择一位专家，并遵循其建议。这需要根据可观测的专业知识上下文，确定哪位专家的建议能够带来最大的预期奖 $\vec{z}_t$励。实际上，我们将问题化简为一个上下文多臂老虎机。具体而言，构建的老虎机中的每个臂都对应原问题中的一位专家，选择该臂就意味着遵循相应专家的建议。我们观测到所选臂的奖励后，便可更新所遵循建议的那位专家的质量估计。反过来，每位专家（即构建的老虎机中的每个臂）的质量取决于专业知识上下文。通过将适当的上下文老虎机算法应用于化简后的问题，我们可以学习近似这一函数，同时确保充分探索。这种方法可以看作是将元多臂老虎机方法（（Auer 等人，2002a），即将每位专家建模为元老虎机中的一个臂，并使用标准多臂老虎机算法来确定最佳专家）推广到上下文情形。

所选的上下文老虎机算法会影响可学习的专长。例如，基于决策树的算法（如 *TreeHeuristic*(Elmachtoub 等人，2017年）更适用于离散的专业知识区域。基于树的算法还有一项优势：它们会执行特征选择；如果专业知识上下文包含许多特征，而其中只有少数能够预测专长。

这种化简的一项重大缺点是，决策仅依据单个专家的建议作出。除了将每轮的知识获取限制为只能依赖一位专家——当专家数量较多时，这通常会导致性能不佳——学习器也不可能在任何给定的专业知识上下文中超越集体中的最佳个体。相比之下，若根据多位专家的建议来作出决策，就有机会实现集体智慧（即性能优于单个最佳专家）。基于这一目标，我们将在下一节中提出：（i）一种受局部学习启发的最近邻方法，并指出其缺点；随后（ii）通过我们新颖的专业知识树方法解决这些缺点。

3. 方法

我们的目标是学习从专家建议到臂的映射，使我们的决策能够最大化累积奖励。由于专家质量是专业知识上下文的函数 $\vec{z}_t$，我们的目标是学习一个以 $\vec{z}_t$ 为条件的映射，也就是说，我们希望学习一个策略 π： $[0,1]^N \times \mathbb{R}^{g\times K} \to \Delta^K$，它将建议和专业知识上下文都映射为臂上的一个分布。我们提出通过将专业知识空间划分为多个区域，并为每个区域分别维护一个独立的“带专家建议的多臂老虎机”学习器实例来学习 π，例如 EXP4（Auer 等人，2002a）或 Meta-CMAB（Abels 等人，2023年）。

如果已有先验知识表明 $\vec{z}_t$ 的哪些特征可能会改变专家的性能，那么就可以利用这些知识对空间进行适当的划分，然后在每个子集上分别应用独立的专家建议多臂老虎机算法。如果没有此类知识，或者高估或低估了特定特征的影响，那么所选择的划分将反映这些不准确之处。当没有关于 appro-

如果事先无法获得适当的划分，一种合理的解决方案是根据以往类似的经验采取行动。

因此，我们考虑一种算法，将局部学习的原理（Bottou &Vapnik，1992年）应用于专家建议多臂老虎机设定。给定一个专业知识上下文$\vec{z}_t$，在时间 t，我们可以对之前的经验进行采样，选取专业知识上下文最接近的 p 比例。这些$\lceil \frac{p}{100} \cdot t \rceil$ 经验随后用于初始化一个专家建议多臂老虎机学习器。我们将这种方法称为最近$p\%$。虽然此方法不需要知道预期划分数量，但应合理选择 p，以避免过拟合或欠拟合。或者，也可以使用距离阈值，而非邻居数量阈值。在这两种情况下，阈值过大都会纳入来自不同区域的先前经验，而取值过小则可能导致模型欠拟合。与阈值选择同样重要的是，该方法的性能取决于距离度量作为模型相似性代理的准确性。具体而言，如果专业知识上下文包含与专家质量无关的特征，那么将其纳入距离计算会影响学习所得模型的质量。为应对这些不相关特征并自动学习恰当的邻域，我们将在下一节提出一种新方法，通过决策树同时学习识别相关特征和划分。

3.1. 用专业知识树捕捉局部知识

模型树（Wang& Witten，1996年）是一种决策树，在其叶节点上拟合回归模型（通常为线性模型）。这种决策树能够有效地将空间划分为多个区域，使得在每个区域内使用单独的模型更有利。我们在此开发一种模型树算法，使其适用于具有局部专家建议的多臂老虎机场景。这种场景与通常的模型树场景有两个方面的不同。首先，多臂老虎机场景需要探索，并且优化目标不同（关注奖励而非预测误差）。其次，对于传统模型树，划分空间所依据的特征与叶节点模型所使用的特征相同。与模型树不同，我们划分的是专业知识空间（即，依据 $\vec{z}_t$），并基于专家建议学习模型（即，$\boldsymbol{\xi}_t$）。鉴于这一差异，我们将其称为专业知识树。

3.1.1. 构建专业知识树

设 $\mathcal{H}$ 为由经验元组构成的经验历史，其中包含建议矩阵、所选动作、收集到的奖励、该动作的概率以及专业知识上下文，即一个 $\langle \boldsymbol{\xi}, k, r, p, \vec{z} \rangle$ 元组。给定 $\mathcal{H}$，我们可以沿着专业知识上下文的 i^{th} 维度对这组经验进行划分，从而得到两个子集，分别记为$\mathcal{H}_{<\tau^{(i)}}$：
$= \{\langle \boldsymbol{\xi}, k, r, p, \vec{z} \rangle \in \mathcal{H} | z^{(i)} < \tau^{(i)}\}$，并且类似地。

算法 1（增量式）专业知识树

要求： N 名专家，奖励函数为
数 f：$[K] \times \mathrm{R} \to [0,1]$
$root \leftarrow$初始化一棵空的专业知识树
对于 $t = 1, 2 ..., T$ **do**
  专家观察上下文 $\vec{x}_t$
  获取 $\vec{x}_t$ 的可观测子集，即专长
  上下文 $\vec{z}_t$
  获取专家建议 $\boldsymbol{\xi}_t = \{ \vec{\xi}^1(\vec{x}_t) ..., \vec{\xi}^N(\vec{x}_t)\}$
  $node \leftarrow root$
  当 $node$ 存在有益划分**时**
    $node \leftarrow$ $node$ 中包含 $\vec{z}_t$的子节点
  结束循环
  令 π_t 为 $node$的学习器所采用的策略
  拉取臂 $k_t \sim \pi_t(\boldsymbol{\xi}_t)$，并收集由此产生的奖励 r_t
  $p \leftarrow \pi_{k_t,t}(\boldsymbol{\xi}_t)$ {获取所选臂的概率}
  如果不增量**那么**
    $node \leftarrow root$ {从根节点开始更新树}
  结束 if
  更新 $node$ 与 $\langle \boldsymbol{\xi}_t, k_t, r_t, p \rangle$
  当 节点不为空**时**
    更新以 $node$ 为根的候选划分 $\langle \boldsymbol{\xi}_t, k_t, r_t, p \rangle$
    如果 最佳候选划分满足公式（1）成立，**则**
      分裂 $node$
    结束如果
    将 $node$ 设为包含 $\vec{z}_t$ 的子节点
  结束 while
结束 for

地以 $\mathcal{H}_{\geq\tau^{(i)}} := \mathcal{H} \setminus \mathcal{H}_{<\tau^{(i)}}$ 进行，其中 $z^{(i)}$ 表示专业知识上下文 $\vec{z}$的第 i个值，而 $\tau^{(i)}$ 是我们进行划分的阈值。随后，这两个经验子集可用于维护两个专家建议多臂老虎机算法，我们分别将其记为 $\mathrm{ALG}_{<\tau^{(i)}}$ 和 $\mathrm{ALG}_{\geq\tau^{(i)}}$。令 $Q(\mathrm{ALG}\ \mathcal{H})$ 表示某个算法 ALG 在经验集 $\mathcal{H}$上的表现估计，具体定义将在下一节中给出。只有当这两个子集学习器的表现优于在 $\mathcal{H}$上训练的单个学习器 ALG 时，我们才能从划分中获益；当以下条件成立时，情况正是如此：

$$Q(\mathrm{ALG}_{<\tau^{(i)}}, \mathcal{H}_{<\tau^{(i)}})|\mathcal{H}_{<\tau^{(i)}}| + Q(\mathrm{ALG}_{\geq\tau^{(i)}}, \mathcal{H}_{\geq\tau^{(i)}})|\mathcal{H}_{\geq\tau^{(i)}}| > Q(\mathrm{ALG}, \mathcal{H})|\mathcal{H}| \quad (1)$$

请注意，这两个质量估计分别根据其对应子集的大小进行加权。

3.1.2. 估计专长质量

给定一组经验 $\mathcal{H}$，一个值得探讨的问题是：另一种策略在平均奖励方面会有怎样的表现。按照Auer 等人 [2002b], ，我们可以

估计学习器 ALG 及其策略 π^{ALG} 在这组经验上的期望平均奖励，如下所示

$$Q(\mathrm{ALG}, \mathcal{H}) = \frac{1}{|\mathcal{H}|} \sum_{\langle \boldsymbol{\xi}, k, r, p, \vec{z} \rangle \in \mathcal{H}} \pi_k^{\mathrm{ALG}}(\boldsymbol{\xi}) r / p \quad (2)$$

其中，$\pi_k^{\mathrm{ALG}}(\boldsymbol{\xi})$ 表示在给定建议向量 $\boldsymbol{\xi}$ 的情况下，策略采取行动 k 的概率。由于这些经验是由策略获取的（因此存在偏差），分母中的概率可确保期望奖励无偏。

因此，我们考虑以下方法：构建一棵专业知识树，其中用于划分的特征是 $\vec{z}$的特征，然后在叶节点中的经验上运行专家建议多臂老虎机算法。更具体地说，只要存在有益划分（即满足（1）的划分），我们就反复将专业知识空间划分为若干子集，并为这些划分分别维护学习器。在满足（1）的划分中，我们选择使不等式左侧最大化的划分。算法1给出了伪代码。这种方法有效地将专家建议多臂老虎机算法与决策树的灵活性结合起来。最终得到一个两步模型：一方面是关于 $\vec{z}$ 的决策树，另一方面是每个叶节点中的模型。这有效地将$\vec{z}$空间划分为多个区域，每个区域都有自己的策略。

由于每个潜在分裂都需要模拟两个专家建议多臂老虎机算法的运行，因此考虑连续变量的所有可能分裂会变得难以处理。相反，遵循 Potts & Sammut [2005]，的做法，我们为每个特征固定 κ 个分裂候选项。虽然这会稍微降低灵活性，但也能大幅减少计算时间。

3.1.3. 增量专业知识树

由于每次获得经验后都重新训练专业知识树的成本很高，我们提出一种增量构建树的算法：也就是说，在每个时间步，根据收集到的经验更新模型，并在能够提升专业知识树质量时进行分裂。与非增量变体不同，分裂一旦完成便不会重新审视。因此，如果问题的随机性导致了不佳的分裂，就可能影响之后的性能。请注意，质量估计值（2）的方差很大，因此，尝试通过例如使用 Hoeffding 界（Pfahringeretal., 2007）来限制早期分裂的启发式方法作用有限。不过，分裂带来的额外探索（如第 4节所述）可以抑制不必要的分裂。

这种权衡是性能可能有所下降，但算法的效率显著提高。完整学习一棵专业知识树比增量更新一棵专业知识树更耗时。更具体地说，假设树的理想深度为 h（而这又与划分数量的对数成正比），专业知识上下文的维度为 g，且每个特征有 κ 个候选划分，那么构建相应的专业知识树需要估计 $O(2^h g\kappa)$ 个模型。更新专业知识树需要进行 $O(hg\kappa)$ 次更新（对于专业知识上下文所在叶节点路径上的每个节点，都要进行 $g\kappa$ 次更新）。相比之下，更新增量树平均只需进行 $O(g\kappa)$ 次模型更新，因为只更新叶节点模型。

4. 理论结果

下面，我们给出局部专长问题的一些界，并阐明不恰当地划分上下文空间所付出的代价。

令 $\mathcal{Z}$ 为专长区域，即 R^g 的一个划分，其中每个子集内的专长都是同质的；换言之，每个子集都对应一个唯一的最优策略。

局部专长的代价 我们首先为具有局部专长的多臂老虎机算法的遗憾建立一个下界。假设已知 $\mathcal{Z}$，因此我们可以运行 $|\mathcal{Z}|$ 个带专家建议的多臂老虎机算法实例，并将每次具有专业知识上下文 $\vec{z}$ 的经验分配给相应的学习器。记 $p(Z)$ 为抽取的上下文属于区域 $Z \in \mathcal{Z}$ 的概率。令 $R(T)$ 表示一个带专家建议的多臂老虎机算法实例所产生的遗憾。所有实例产生的总遗憾为 $\sum_{Z\in\mathcal{Z}} R(p(Z)T)$。在通常情况下，当 $R(T) \propto \sqrt{T}$ 时，该量在 p 的熵达到最大时取最大值；$p(Z) = 1/|\mathcal{Z}|\ \forall Z \in \mathcal{Z}$，从而导致遗憾为 $|\mathcal{Z}|R(T/|\mathcal{Z}|)$。反之，当 $\max_{Z\in\mathcal{Z}} p(Z) = 1$ 时，该量取最小值；这等同于不存在局部专长，因此遗憾降至 $R(T)$。

非局部化算法的预期遗憾 对于任意专家 n，令 $p'(n)$ 表示专家 n 为最优的专业知识上下文所占的比例。由此可知，任何非专业化算法，即无论上下文如何都选择同一专家的算法，在每个区域中选择最佳专家的概率至多为 $p'(n^*) = \max_{n\in[N]} p'(n)$。特别地，对于专家为最优的上下文，该算法产生的遗憾为 0；而对于所有其他上下文，则会产生恒定遗憾。因此，遗憾的相应下界为 $\Omega((1 - p'(n^*))T)$。因此，在局部情境中，任何非专业化算法都必然产生线性遗憾。当 p' 的熵达到最大时，该下界最大；在此情况下，产生的遗憾为 $\Omega(\frac{N-1}{N}T)$。

（不）进行拆分的成本令 $\mathcal{H}$ 为待拆分的经验集合，令 $\tau^{(i)}$ 为 i特征上的阈值。令 $\mathcal{Q} = |\mathcal{H}|Q(ALG, \mathcal{H})$ 表示学习器在具有历史记录 $\mathcal{H}$ 的父节点上的期望累积奖励，如公式 2所定义。同样地，我们分别将左子节点和右子节点上的性能记为 $\mathcal{Q}_l = |\mathcal{H}_{<\tau^{(i)}}|Q(ALG, \mathcal{H}_{<\tau^{(i)}})$

以及 $\mathcal{Q}_r = |\mathcal{H}_{\geq\tau(i)}|Q(ALG,\ \mathcal{H}_{\geq\tau(i)})$。

当 $\mathcal{Q} < \mathcal{Q}_l + \mathcal{Q}_r$ 时，有必要进行分裂。换句话说，如果保留两种策略能够提升性能。

设 p 为：在上下文包含于父节点的前提下，该上下文包含于左子节点的概率。相应地，一段经验包含于右子节点的概率为 $1-p$。因此，分裂的预期遗憾为 $R(pT) + R((1-p)T)$。其中，pT和 $(1-p)T$分别为分配给左、右子节点的经验数量期望值。

未分裂节点上的性能可以表示为未分裂节点上的预期累积奖励与分裂后预期累积奖励的加权和之间差值的函数。

具体而言，令 $\mathcal{R} = \max_\pi \mathrm{E}[f(k_{\pi(\boldsymbol{\xi}(\vec{x}))}\vec{x})],$，其中期望是针对从上下文空间相关区域采样的上下文计算的，它表示父节点所含经验上最佳策略的期望奖励。类似地，令 $\mathcal{R}_l$ 和 $\mathcal{R}_r$ 分别表示左子节点和右子节点上的最大期望奖励。

接着，令 $\Delta = p\mathcal{R}_l + (1-p)\mathcal{R}_r - \mathcal{R}$ 表示不分裂与分裂学习器之间期望平均奖励的差距。根据遗憾的定义，取期望后可知，学习器在父节点经验上的累积奖励为 $\mathrm{E}[\mathcal{Q}] = \mathcal{R}T - R(T)$，即最优性能减去遗憾。类似地，在子节点经验上的期望性能为 $\mathrm{E}[\mathcal{Q}_l] = \mathcal{R}_l Tp - R(Tp)$，且 $\mathrm{E}[\mathcal{Q}_r] = \mathcal{R}_i T(1-p) - R(T(1-p))$

然后，我们可以将一次划分的预期改进表示为：

$$\begin{aligned}
&\mathbb{E}[\mathcal{Q}_l + \mathcal{Q}_r - \mathcal{Q}] \\
&= (\mathcal{R}_l p + \mathcal{R}_i(1-p) - \mathcal{R})T - R(Tp) - R(T(1-p)) \\
&\quad + R(T) \\
&= \Delta T - (R(Tp) + R(T(1-p)) - R(T)) \\
&= \Delta T - (\sqrt{p} + \sqrt{1-p} - 1)R(T)
\end{aligned}$$

其中，最后一个等式假设 $R(T) \propto \sqrt{T}$（这对于专家建议算法而言是典型情况，例如 EXP4- IX（Neu，2015年）或 Meta-CMAB（Abelsetal.，2023年））。

具体而言，这意味着只有当性能差距足够大时，分裂才有益。这是分裂所引入的额外探索造成的后果。因此，如果专家之间足够接近，即如果$\Delta < (\sqrt{p} + \sqrt{1-p} - 1)R(T)/T$，分裂反而会造成损害。由于$R(T)/T$ 通常随臂的数量 K 和专家的数量 N单调增加，但随时间步的数量 T单调减少，这印证了这样的直觉：当T较大，或当 K 和 N （以及它们引入的探索成本）较小时，分裂最有益。

其中一个特殊情况是 $\Delta = 0$，即未分裂节点中的最优策略与子节点中的策略完全相同。在这种情况下，所有学习器都会趋向于相同的策略，但子节点会进行一些额外的探索，从而导致性能下降。具体而言，这一额外的探索惩罚为$R(Tp) + R(T(1-p)) - R(T)$。由于合理的专家建议多臂老虎机算法对 T呈现次线性依赖关系，所得总和为正。因此，分裂会带来与 $\sqrt{pT} + \sqrt{(1-p)T} = (\sqrt{p} + \sqrt{1-p})\sqrt{T}$成比例的遗憾。这是父节点遗憾的（$\sqrt{p} + \sqrt{1-p}$）倍。在这种情况下，当 $p = 0.5$时，该总和最大，此时不必要的分裂会使遗憾放大 $\sqrt{2}$倍。

5. 实验设置

我们根据平均奖励评估我们的算法。在全文中，我们比较以下算法的性能：*Meta-CMAB*，一种针对具有（非局部化）专家建议的多臂老虎机的先进方法；*TreeHeuristic*（Elmachtoub 等人，2017年），将其应用于我们在第 2.2 节中提出的归约；*Oracle*，该算法依赖于最优划分的先验知识，并为每个区域维护一个独立的 Meta-CMAB 实例；*Nearest*1% 和 10%，它们分别基于使用过去经验中最近的1% 或 10% 训练的模型做出决策；最后是第 （增量式）专业知识树算法，如第 3.1 节所述。我们的结果取 100 次模拟的平均值，并针对不同数量的臂和专家进行实验。用于复现这些结果的代码已在补充材料中提供。

5.1. 多臂老虎机问题

遵循Riquelme 等人 [2018], 的方法，我们将分类问题构建为多臂老虎机。因此，给定一个上下文，单个臂（对应真实标签）会提供1的奖励，而其他臂则提供0奖励。我们从 OpenML 数据仓库中选取了多个数据集进行评估，这些数据集在特征分布、臂的数量以及各臂上的奖励分布方面各不相同（Vanschoren 等人，2014）。数据集的选择和处理方法详见补充材料。

5.2. 模拟局部专长

我们研究专长的变化，并通过专长热图来刻画这些变化；热图将专业知识上下文映射到专家质量。每轮中，专家会针对一个上下文向量提供建议 $\vec{x}_t$，其中的一部分 g 特征（在实验开始时随机选取）构成专业知识上下文 $\vec{z}_t$。在这些g 特征中，2 个特征与专长相关，如图 2所示。因此，合适的学习器必须学会哪些 g 特征是相关的，同时还要

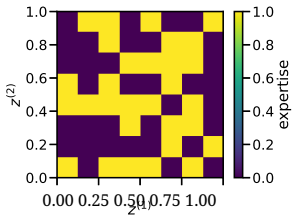

图 2：具有 64 个区域的专长热图示例。热图通过随机为每个单元格分配 0 或 1 的专长值生成。每个时间步都会从该空间中采样一个点，并根据每位专家各自的热图确定其专长。专长值较高时，专家会提供诚实建议；专长值较低时，则会提供对抗性建议。

学会进行适当的划分。每位专家都有一张独立的热图；对于给定的 $\vec{z}_t$ 值，该专家的质量由相应的专长值决定。接近 0 的值会导致对抗性建议，而较高的专长值则会带来诚实建议。热图的生成方式是：为每位专家在各个区域（区域数为 1、4、16 或 64）分别赋予 0 或 1 的专长值。区域数越多，问题就越难。例如，在我们考虑的最难配置中，要优化性能，就必须根据有限的经验正确识别 64 个区域。通过增大 g 的大小，我们增加不相关特征的数量，从而使问题更加复杂。

6. 结果与讨论

我们以平均奖励而非遗憾来展示结果，因为这样既能呈现区域数量增加所造成的损失，也能展示各算法性能的损失。若以遗憾来呈现结果，前一项信息就会被掩盖，因为 Oracle 和各个算法都会遭受这种损失。对于所有专家数量（$N \in \{4, 32\}$）与数据集的组合，结果均取 $T = 1000$ 个时间步的平均值。

6.1. 专长树算法对问题复杂度的增加更具鲁棒性

在这种设定下进行学习具有两方面的挑战：既必须识别出合适的区域，也必须为每个区域学习合适的模型。为单独考察因需要识别这些区域而导致的性能损失，我们将自己的算法与 Oracle 算法进行比较。对于该算法，我们假设已完全了解这些区域，因此 Oracle 变体的任何性能损失都仅源于对每个区域适用模型的不确定性。这使我们能够单独考察非预言机变体的性能损失有多少是由对专业知识空间结构的不确定性所致。

图 3展示了随着区域数量增加，我们的算法性能所发生的变化。这些图表表明，当存在专业化（即区域数量大于1 个）时，非专门化的 Meta-CMAB 算法无法进行学习。这些图表还表明，随着区域数量的增加，专门化算法的性能下降速度快于 Oracle。这是由于对专业知识上下文空间划分的不确定性所致。这是因为，要为数量较多的区域找到合适的划分，需要进行多次最优分裂。图 3 还表明，对于给定数量的区域，专业知识上下文的规模越大，就越难以确定理想的划分，这一点同样可以从与 Oracle 性能之间不断扩大的差距中看出。这是因为，当特征数量较多时，可能出现随机相关性，从而导致产生不具备实际区分能力的分裂。尽管这会导致所有专门化算法的性能下降，但对这两种专业知识树变体的影响较小。ExpertiseTree 的增量变体受不相关特征的影响更大，因为由随机相关性产生的分裂无法撤销。

在 1% 或 10% 的经验数据上训练的最近邻模型，分别假设专业知识空间大约划分为 100 个或 10 个区域。当区域数量逐渐趋近前者时，1% 版本的性能相较于 10% 版本有所提升。此外，随着专业知识上下文规模的增大，距离度量作为专家质量相似性代理指标的质量会下降。这会导致最近邻算法的性能下滑，即使区域数量较少也是如此。

最后，尽管 TreeHeuristic 算法可以像 ExpertiseTree 算法一样进行特征选择和划分，但其单专家方法会使其处于劣势。这不仅会影响 TreeHeuristic 在任意时间步所能做出的最佳决策，也会影响其更新，因为每条经验都只用于更新一位专家的质量估计。这实际上意味着，与我们的方法相比，TreeHeuristic 的每位专家所拥有的经验少了 N。

请注意，对于所有算法，包括 Oracle，相较于非局部化设置，性能下降是不可避免的。区域越多，需要的单个模型就越多；但由于经验数量固定，每个单个模型的训练次数都会减少，导致性能变差。

6.2. 基于全局专业知识的局部化算法

如果专长并未局部化，局部化模型的性能可能会更差，因为它们考虑的相关经验更少。图3 中每个子图最左侧的点展示了专家未局部化（即只有一个区域）时各种算法的性能。值得注意的是，TreeHeuristic 的性能明显

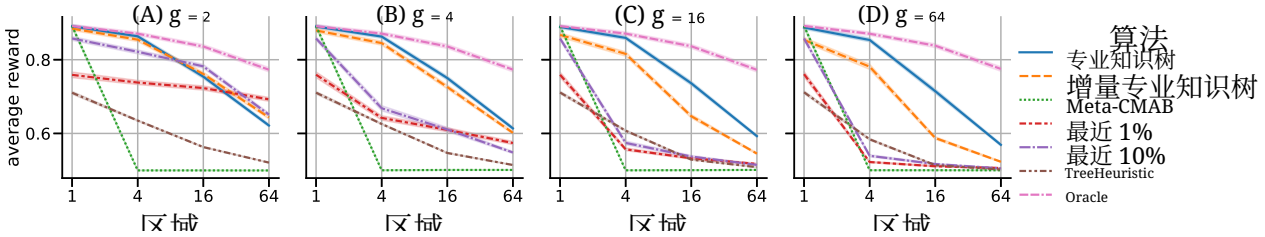

图3：不同算法在不同区域数量（横轴）和专业知识上下文大小（g，列）下的平均奖励。阴影区域表示95%的置信区间。

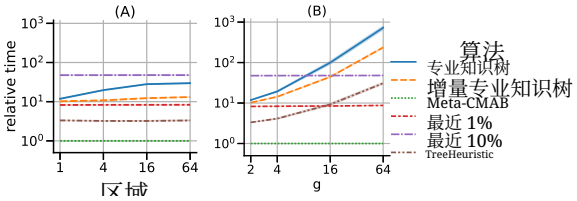

图4：不同数量的（A）区域或（B）专业知识上下文大小下的相对执行时间。相对时间通过算法的执行时间除以 Meta-CMAB 的执行时间得出。两个坐标轴均为对数刻度。ExpertiseTree方法的执行时间随区域数量呈对数增长，随专业知识上下文大小呈准线性增长。

逊于其他算法。这种性能损失源于 TreeHeuristic 的策略：它会为每个学习得到的划分选取一位专家。因此，其性能上限取决于表现最佳的专家。这是其他算法所没有的局限。不过值得注意的是，与非专门化的 Meta-CMAB 相比，最近邻算法的性能也有所下降。这些算法基于全部经验的一个子集学习模型，因此学习到的模型泛化能力较弱，导致性能下降。当纳入的经验仅占 1% 时，这种情况尤为明显。相比之下，由于不存在有益划分，ExpertiseTree方法会保留单一划分。即使专业知识上下文的规模显著增大，ExpertiseTree 的性能仍与简单的 Meta-CMAB 几乎相同。随着专业知识上下文规模增大，增量 ExpertiseTree 的性能损失则更为明显。规模越大，越容易观察到虚假相关性，从而导致不必要的分裂，而增量 ExpertiseTree 无法撤销这些分裂。

6.3. 执行时间

图 4 表明，大多数算法的执行时间不会随区域数量或专业知识上下文大小而变化。具体而言，最近邻算法在每一轮仅训练一个模型，所使用的经验集不受区域数量或专业知识上下文大小的影响。然而，邻域大小确实有显著影响。邻域越大，模型训练所用的经验就越多，因此执行时间也越长。请注意，专业知识上下文大小增大会增加距离计算的成本，从而使执行时间略有增加。

ExpertiseTree 在平均奖励方面取得提升的代价是计算时间显著增加。这些图证实了执行时间在理论上的增长：与区域数量呈对数关系，与专业知识上下文的规模呈线性关系。由此产生的执行时间对于时间敏感型问题而言可能过长，但对于例如涉及人类专业知识的问题来说仍属合理。请注意，尤其是 ExpertiseTree 的成本取决于潜在分裂的数量，因此取决于专业知识上下文的规模。通过剔除不相关特征（可以预先剔除，也可以依据观测数据剔除），能够降低这一成本。

7. 结论

在本研究中，我们探讨了专家建议多臂老虎机的一种推广形式，该形式对局部专长进行了显式建模。在现实世界的场景中，人类和人工智能的专长取决于一组特征，因此，能够考虑局部专长的方法至关重要。鉴于以往在该场景中的算法并未以这组特征为条件，我们提出了两种性能和复杂度逐步提升的新算法。第一种算法依赖最近邻查询来识别相关经验，但当邻域大小不合适，或距离度量无法反映专家质量的相似性时，该算法会失效。为解决这些缺陷，我们的专业知识树方法学习一棵树，根据相关特征对专业知识空间进行划分，然后学习在每个学习得到的划分中利用所有专家建议采取行动。即使区域数量或专业知识上下文的规模较大，该方法也能保持良好的性能。为了增强其适用性，我们认为未来的工作应解决 ExpertiseTree 的高计算成本问题。特别是，在线特征选择可能有助于降低高维上下文所带来的计算成本。

致谢

A.A. 获得了比利时国家科学研究基金（F.N.R.S.）提供的 FRIA 资助（编号：5200122F ）。T.L. 获得了 F.N.R.S.项目的资助，资助编号为 31257234 和 40007793；还获得了 F.W.O.项目的资助，资助编号为 G.0391.13N；以及 DigitalWallonia4.ai 通过瓦隆公共服务研究部提供的资助，资助编号为 2010235–ARIAC。T.L. 和 A.N. 获得了佛兰德政府通过人工智能研究计划提供的支持。T.L.、V.T. 和 A.N. 感谢 TAILOR 项目的支持；该项目由欧盟“地平线2020”研究与创新计划资助，GA编号为 952215。本研究使用的资源和服务由 VSC（佛兰德超级计算中心）提供，该中心由佛兰德研究基金会（FWO）和佛兰德政府资助。

参考文献

Abels, A., Lenaerts, T., Trianni, V., and Now´e, A. 应对集体决策中的专家偏差。人工智能，第 103921 页，2023年。Auer, P., Cesa-Bianchi, N., and Fischer, P. 多臂老虎机问题的有限时间分析。机器学习，47(2-3):235–256，2002a。Auer, P., Cesa-Bianchi, N., Freund, Y., and Schapire, R. 非随机多臂老虎机问题。《*SIAM*计算杂志》，32(1):48–77，2002b。doi: 10. 1137/S0097539701398375。Bond, S. D., Carlson, K. A., Meloy, M. G., Russo, J. E., and Tanner, R. J. 单一选项评估中的信息扭曲。《组织行为与人类决策过程》，102(2):240–254，2007。Bottou, L. and Vapnik, V. 局部学习算法。神经计算，4(6):888–900，1992年。Chu, W., Li, L., Reyzin, L., and Schapire, R. 具有线性收益函数的上下文老虎机。载于第十四届人工智能与统计国际会议论文集，第 208–214 页，2011。Elmachtoub, A. N., McNellis, R., Oh, S., and Petrik, M. 一种使用决策树求解上下文老虎机问题的实用方法。*CoRR*，abs/1706.04687，2017年。Foster, D. and Rakhlin, A. 超越置信上界：利用回归预言机实现最优且高效的上下文老虎机。载于国际机器学习会议，第 3199–3210 页。PMLR，2020。Gianfrancesco, M. A., Tamang, S., Yazdany, J., and Schma- juk, G. 使用电子健康记录数据的机器学习算法中可能存在的偏差。《*JAMA*内科学》，178(11):1544–1547，2018年。Neu, G. 不要再探索：改进非随机多臂老虎机的高概率遗憾界。神经信息处理系统进展，28，2015年。Pfahringer, B.、Holmes, G. 和 Kirkby, R. Hoeffding 树的新选项。载于澳大利亚人工智能联合会议，第 90–99 页。Springer，2007 年。Pohl, R. 认知错觉：关于思维、判断和记忆中的谬误与偏差的手册。2017 年。Potts, D. 和 Sammut, C. 线性模型树的增量学习。机器学习，61(1-3)：5–48，2005 年。Riquelme, C.、Tucker, G. 和 Snoek, J. 深度贝叶斯多臂老虎机对决：用于 Thompson 抽样的贝叶斯深度网络的实证比较。载于国际学习表征会议，2018 年。Thompson, W. R. 关于根据两组样本的证据判断一个未知概率超过另一个未知概率的可能性。*Biometrika*，25(3/4)：285–294，1933 年。doi：10.2307/2332286。Trope, Y.、Gervey, B. 和 Liberman, N. 从实用主义假设检验视角看一厢情愿的想法。神话狂想：欺骗与自我欺骗的本质，第 105–31 页，1997 年。Valko, M.、Korda, N.、Munos, R.、Flaounas, I. 和 Cris-tianini, N. 核化上下文老虎机的有限时间分析。载于第二十九届人工智能不确定性会议论文集，第 654–663 页，2013 年。Vanschoren, J.、Van Rijn, J. N.、Bischl, B. 和 Torgo, L. Openml：机器学习中的网络化科学。*ACM SIGKDD* 探索通讯，15(2)：49–60，2014。Wang, Y. 和 Witten, I. H. 用于预测连续类别的模型树归纳。1996年。Whitehead, S. J. 和 Ali, S. 经济学评估中的健康结果：质量调整生命年与效用。英国医学公报，96(1)：5–21，2010 年。

附录

A. 数据集选择

我们从 openmpl（Vanschoren 等人，2014）存储库中抽取数据集，这些数据集符合以下标准。

1. 足够数量的实例：为了能够在每个数据集上重复我们的实验，我们将实验限制在至少包含 10，000 个实例的数据集上。

2. 分类：我们将实验限制在分类多臂老虎机上。3. 缺失值：我们仅考虑完整数据集，因此会舍弃包含缺失值的数据集。

这些筛选条件最终选出了以下数据集：mushroom、adult、letter、nursery、pendigits、BNG(page- blocks,nominal,295245)、BNG(glass,nominal,137781)、BNG(tic-tac-toe)、BNG(vote)、electricity、covertype、kropt、BNG(breast-w)、BNG(page-blocks)、BNG(glass)、mammog- raphy、眼动、mozilla4、KDDCup99、MagicTele- scope、Click prediction s mall、artificial-characters、bank- marketing、eeg-eye-state、kr-vs-k、ldpa、skin-segmentation、spoken-arabic-digit、walking-activity、volcanoes-b1、creditcard、Amazon e mployee access、CreditCardSubset、PhishingWebsites、Diabetes130US、numerai28.6、fars、shuttle、Run or walk information、tamilnadu-electricity、jungle chess 2pcs raw endgame complete、MiniBooNE、jannis、helena、microaggregation2

对于这些数据集中的每一个，我们都会对所有特征进行虚拟变量编码和归一化。