

arXiv:2206.00423v2 [cs.LG] 2022年8月9日

开放环境机器学习

周志华

计算机软件新技术国家重点实验室
南京大学, 南京 210023
zhouzh@nju.edu.cn

抽象的

传统的机器学习研究通常假设近距离环境场景

学习过程中的重要因素保持不变。随着数学的巨大成功，

汉语学习，如今，越来越多的实际任务，特别是涉及

开放环境重要因素可能发生变化的情景，称为打开-

环境机器学习（本文介绍的 Open ML 都是向社区提出的。

显然，从近距离环境转向

开放环境。这变得更加具有挑战性，因为在各种大数据任务中，数据

通常会随着时间的流逝而积累，例如流训练机器很困难

像传统研究一样，在收集所有数据后建立学习模型。本文简要介绍了

介绍了该研究领域的一些进展，重点关注新兴技术

新的类别、递减/递增特征、变化的数据分布、多样化的学习

目标，并讨论了一些理论问题。

1. 简介

机器学习在各种应用领域取得了巨大成功，特别是在极好的-

有远见的学习分类和回归等任务。在机器学习中，通常

预测模型优化一个特定的目标，是从训练数据集中学习而来的，

训练样例，每个样例对应一个事件/对象。训练样例

由两部分组成：特征向量（或称为实例）描述外观

事件/对象，以及标签表示相应的真实输出。在分类
在回归分析中，标签表示训练实例所属的类；在回归分析中，
label 是与实例对应的实值响应。本文主要关注
分类，尽管大多数讨论也适用于回归和其他机器
学习任务。正式地，考虑学习任务 $f: X \rightarrow Y$ 来自训练数据集
 $D = \{(x_1, y_1), \dots, (x_m, y_m)\}$ ，在哪里 $x_i \in X$ 是特征空间中的特征向量 x ，和
是 $y_i \in Y$ 是给定标签集中的真实标签是。

值得注意的是，目前机器学习的成功主要体现在以下任务上：关闭-
环境场景，其中学习过程的重要因素保持不变。对于
例如，所有要预测的类标签都是预先知道的，描述
训练/测试数据永远不会改变，所有数据都来自相同的分布，并且
学习过程朝着一个不变的唯一目标进行优化。图 1 说明了
封闭环境机器学习研究中假设的那些典型不变因素。

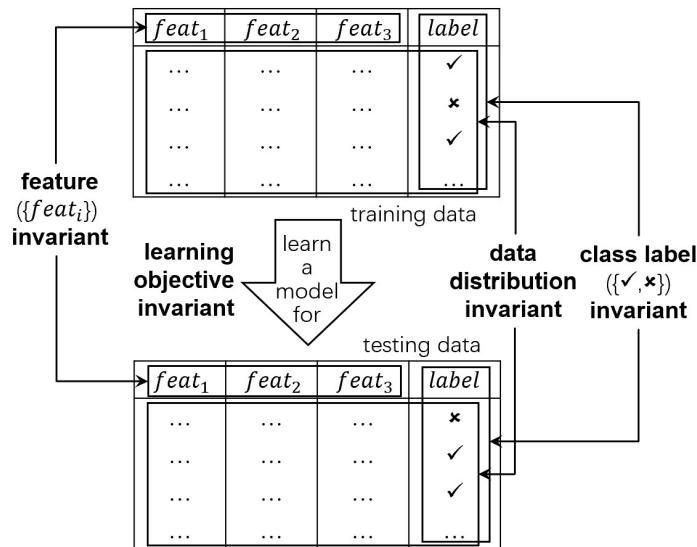


图 1：封闭环境机器学习中假设的典型不变因素。

近环境假设提供了一种简化的抽象，使得复杂的
使任务能够以更简单的方式处理，从而推动机器的蓬勃发展
学习技术。随着这些技术取得的巨大成就，如今，
越来越多的挑战性任务超出了近距离环境的设定。

社区，要求新一代机器学习技术能够
处理更改学习过程中的重要因素。我们称之为开放环境
机器学习¹或简而言之，开放式学习或者打开机器学习。

似乎有一个简单的解决方案：人工生成许多训练示例
通过提前模拟可能的变化，然后将这些数据提供给强大的
机器学习模型，如深度神经网络。然而，这样的解决方案只是
适用于用户了解或至少可以估计哪些变化和
变化将如何发生。不幸的是，在大多数实际任务中情况并非如此。
当我们考虑到现实大数据任务中的数据时，挑战就更大了
通常会随着时间推移而累积，例如，实例会一个接一个地被接收，就像
溪流在我们获得机器学习模型后，我们不可能再对其进行训练所有数据手头上有
在传统研究中，更合理的方法是让训练好的模型
根据新收到的数据进行改进/更新。不幸的是，众所周知
那灾难性遗忘[59] 如果要完善已训练的深度神经网络，则可能会发生
仅使用新数据，而基于存储所有接收到的数据的频繁重新训练可能
导致难以承受的巨大计算和存储成本。尽管有研究
持续学习[13] 尝试帮助深度神经网络抵抗遗忘，经过多次尝试
通常需要扫描大批量训练数据并进行离线训练，
大流数据的严重计算和存储问题。

尽管面临巨大的挑战，但最近在 Open ML 方面仍投入了大量的研究精力。
本文将简要介绍该研究领域的一些进展，重点介绍技术
关于新兴类别、递减/递增特征、变化数据的独特之处
分布和不同的学习目标。一些理论问题也将讨论。

¹ “开放世界机器学习”这个名字曾被用来指代具有看不见的类别的机器学习[57]
或分布外（OOD）数据[61]。事实上，如果看不见的
班级是已知提前，并与第 2 节相关，如果未见过的课程是未知OOD 与第 4 节相关，
尽管仅关注不同的分布比随时间变化的分布更简单。

2. 新兴阶层

近环境机器学习研究通常假设任何未见实例必须是预先已知的给定标签集的成员，即 $\hat{y} \in \gamma$ 。不幸的是，这并不总是成立。例如，考虑一个森林病害监测器借助机器学习模型，利用部署在森林。显然，我们很难预先列举出所有可能的类别标签，因为有些森林疾病可能是全新的，比如由入侵昆虫引起的疾病该地区以前从未遇到过的害虫。为了能够处理 $\hat{y} \in$ /是是一个基本的Open ML 的要求。

我们可能会认为，我们可以人工生成一些虚拟训练示例来新的类别，就像对抗性深度神经网络中采用的流行训练技巧一样，有效。这里的困难在于我们很难想象未知的类（称为新阶级（见下文）可能会发生，而训练一个适应的模型所有可能的类别是不可能的，或者昂贵的难以承受的。

从技术上讲，如果所有数据即将出现，尤其是包括未标记的实例预测，那么处理 NewClass 可以看作是一种特殊的半监督学习[89]任务，例如，通过建立对应于为每个已知类绘制最紧密的轮廓，然后针对未标记的实例在所有轮廓之外作为 NewClass 实例 [12]。实际上，NewClass 的分布可以通过将已知类别的分布与未标记类别的分布分离来近似数据[76]。然而，当数据积累时，这些策略并不直接适用随着时间的推移。

考虑以下新兴类别的学习设置。机器学习模型从一些初始训练数据中进行训练，然后部署来处理看不见的立场如流般涌来。对于已知类别的传入实例，训练模型应该能够做出正确的预测。对于未知类别的传入实例，模型应该能够报告遇到了 NewClass 实例；然后用户可以为 NewClass 创建一个新标签。在遇到此 NewClass 的几个实例后，

训练好的模型应该能够被改进/更新，使得新类别变成一个已知类，其传入实例可以准确预测。理想情况下，整个过程不需要基于存储进行再培训所有数据已收到，因为这在实际的大数据任务中会非常昂贵，甚至不可行。显然，以上描述了一个由人类参与的无监督/监督混合任务。

乍一看，与新兴阶层一起学习似乎与零样本学习，图像分类中的热门话题，旨在对未发生的视觉类别进行分类在训练数据集 [63, 71, 8] 中。请注意，零样本学习假设适用于边信息即类定义/描述/属性等外部知识，可以帮助将可见和不可见的类别联系起来，因此，它可以被视为一种迁移学习[56]；相反，使用新兴类别进行学习是一种通用机器学习学习环境不假设这种外部知识。换句话说，零样本学习假设看不见的类是已知的，尽管它们在训练中没有出现数据，而新兴阶层的学习正在解决的巨大挑战是未见过的类是未知的。因此，针对新出现的类进行学习的方法可以更加通用，并且可以转换并应用于零样本学习。

分类拒绝选项 [19, 3, 24] 旨在避免不自信的预测，由于假设所有类别都已预先知道，因此很可能是不正确的。开放集识别/分类 [60, 25] 扩展了拒绝选项，以考虑未知类别可能出现在测试阶段，目标是识别已知类并拒绝新类。他们试图使训练好的模型适应 NewClass。一些广义的开放集识别研究试图识别未知类别，通过假设可用性零样本学习中前面提到的辅助信息 [25]，而使用新兴类别是一个通用的机器学习设置，它不假设这样的外部知识。

与新兴班级一起学习实际上是一种渐进学习，强调经过训练的模型只需要稍微修改就可以适应新的信息关于增量学习的研究历史悠久[67, 66, 26, 31]，主要是关注训练示例的增量，即E-IL（示例增量学习）

定义见[90]。除了E-IL，[90]中定义的另外两种增量学习分别是A-IL(属性增量学习)和C-IL(类增量学习)。A-IL关于特征增量的cerns，与本文第3部分将要讨论的内容相关，尽管以前的研究通常致力于选择足够的特征空间数据/特征[55, 86]。C-IL关注类增量，与学习有关新兴阶层的出现，尽管先前的研究很少关注NewClass，并且通常假设增量类是已知的[48]。

类发现[27, 51]试图发现稀有类别，将其作为一个独立于类别预测的过程。tion。如上所述，使用新兴类别进行学习是一种无监督/监督的混合任务，虽然这些研究与其第一阶段有一定相关性，但大多没有被侵犯的部分。

在针对新兴类别学习的通用解决方案[52]中，第一阶段是NewClass识别是通过异常检测来实现的。这里的挑战是区分来自已知类异常的新类数据。一般来说，这并不总是可能的；对于例如，图2(a)说明了NewClass和已知异常很难区分类别。幸运的是，在许多实际任务中，假设NewClass实例比已知类的异常更“异常”，作为如图2(b)所示。如果这在原始特征空间中不成立，我们可以尝试通过核映射或表示学习来识别合适的特征空间。之后NewClass实例的识别简化为流中的异常检测，可以通过以下方法解决隔离林[45]。

第二阶段的主要挑战是完善/更新训练模型以适应NewClass，而不会牺牲已知类别的性能。对于深度神经网络，需要基于所有数据（或至少基于智能选择的子样本）进行重新训练以避免灾难性的遗忘，这将带来巨大的计算和存储成本。理想的做法是只进行局部细化以适应NewClass，而不是进行全局细化可能会严重影响已知类别的变化。一种解决方案是利用通过增量式方法仅细化涉及NewClass的树叶，实现树/森林模型方式[52]，它甚至不需要任何已知类数据的存储。替代方案包括

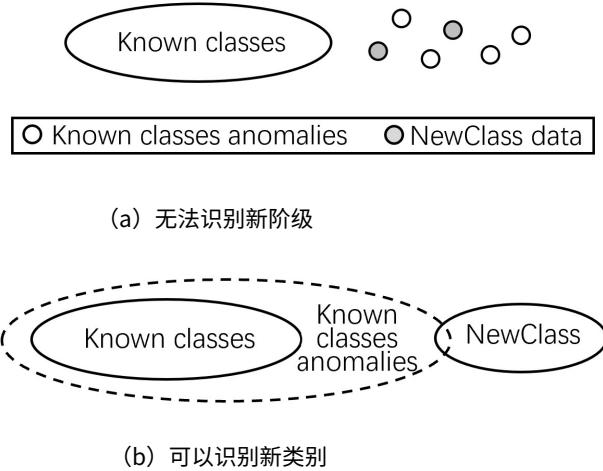


图 2: NewClass 识别并不总是可行的。

可以局部化不同类别影响的技术，使得根据
添加到 NewClass 不会显著影响已知类，例如基于全局的方法
以及局部素描[53]。

如果有多个新类别，则新类别数据的聚类结构可以表示为
被利用[92]。请注意，NewClass
首次检测到时，模型已经完善。为了缩小这一差距，
已经做出了一些努力来使模型能够基于较少的 NewClass 数据进行更新
[93]。多标签学习随着新阶层的出现，挑战性更大，因为
场景中 NewClass 实例可能还包含已知的类标签，甚至可能出现
在已知类别的密集区域中，关键是检测特征的显著变化
组合和/或标签组合[94]。一个相关的主题是检查已知的
类与 NewClass 密切相关，并且评估方法涉及
从新类到已知类的映射已经得到开发[17]。

存在这样的情况：一些 NewClass 实例出现在训练数据中，但被错误地
可能由于特征信息不足而被误认为是已知类实例。这是
更具挑战性的是，目前仅进行了一项非常初步的研究[85]。

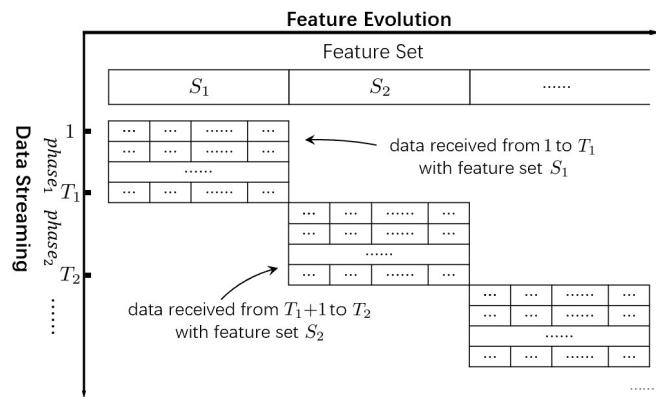
3. 递减/递增功能

近环境机器学习研究通常假设所有可能的实例，
包括看不见的，都位于同一特征空间中，即 $\mathcal{X} \subseteq \mathcal{X}$ 。不幸的是，这确实
并不总是成立。以第 2 节中提到的森林病害监测为例，一些
由于电池耗尽，传感器无法继续发送信号，导致递减
功能，而一些新的传感器可以部署，从而带来增量功能。
能够处理 $\mathcal{X} \setminus \mathcal{X}$ 也是 Open ML 的要求。请注意，与
多样化的阶级，只有新兴阶级才需要特殊待遇，而消失的阶级
可以简单地忽略类，递减和递增特性都需要注意
因为功能减少会导致性能严重下降。

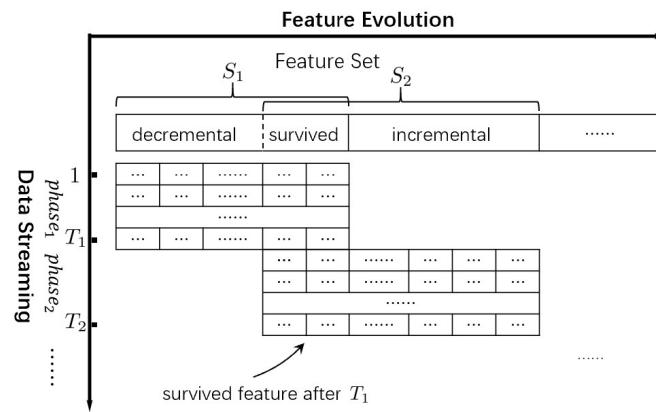
考虑以下具有递减/递增特征的学习设置。
中国学习模型是从一些初始训练数据中训练出来的，然后被部署来处理
看不见的数据像流一样涌来，具有递减和/或递增的特征。对于
对于即将到来的测试数据，模型应该能够做出正确的预测；对于即将到来的
额外的训练数据，模型应该能够相应地进行改进。理想情况下，
希望整个过程不需要基于存储进行重新训练所有数据
已收到。

一般来说，建立一个能够
受益于 $\mathcal{X} \setminus \mathcal{X}$ 为了 $\mathcal{X} \setminus \mathcal{X}$ 因为机器学习是从经验中学习
提高绩效，而在大多数情况下，
学习 $\mathcal{X} \setminus \mathcal{X}$ 什么时候 $\mathcal{X} \cap \mathcal{X} = \emptyset$ 。例如，如图所示
图 3(a)，如果相位特征空间 1 数据（即 $\{(+, y_1), \dots, (+_{\text{电视}1}, y_{\text{电视}1})\}$ ）和相位 2
数据（即 $\{(+, y_{\text{电视}1+1}), \dots, (+_{\text{电视}2}, y_{\text{电视}2})\}$ ）完全不同，那么训练的模型
阶段 1 无助于阶段 2，并且必须基于以下数据从头开始训练新模型
功能集 $\mathcal{X} \setminus \mathcal{X}$ 相 2。

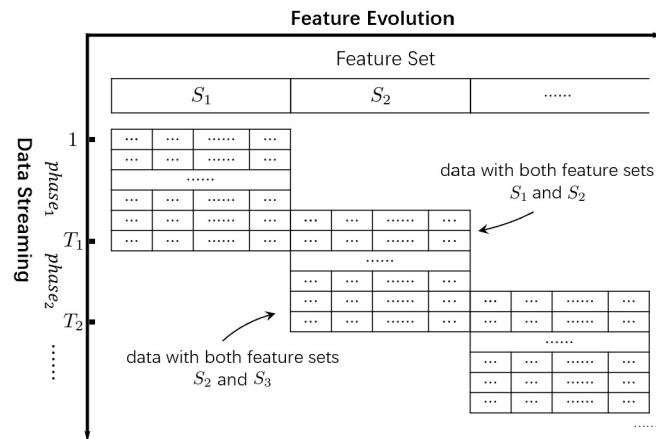
幸运的是，在许多实际任务中， $\mathcal{X} \cap \mathcal{X} = \emptyset$ 换句话说，
是阶段的特征 1 生存在活跃阶段 2 尽管许多其他特征消失了，但
如图 3(b) 所示。例如，不同的传感器可能具有不同的电池寿命，



(a) 阶段1无助阶段2



(b) 阶段1有助于阶段2通过幸存的特征



(c) 阶段1有助于阶段2通过共享实例

图 3：无助/有帮助的特征演变。

因此，在部署新传感器后，一些旧传感器仍在工作。正式地，

同相₁， $\{+\}$ 在 X_s 在哪里 $+$ 和 $+s$ 表示递减和幸存的特征

组分别；同相₂， $\{-\}$ 在 X_s 在哪里 $-$ 表示增量特征集。

作为 $+s$ 除了训练模型之外，在两个阶段中都是共享的从 $+$ ，一个模型₂基于

在 $+s$ 可以分阶段进行训练₁例如[35]：

$$\begin{aligned} \text{分钟} & \sum_{\substack{\text{电视} \\ \text{天}, \text{瓦}_s}}^1 \left(\langle w, x_i \rangle - \langle w, x_i \rangle \right)^2 + \sum_{\substack{\text{电视} \\ \text{我}=1}}^1 \left(\langle \bar{w}_s, x_i \rangle - \langle \bar{w}_s, x_i \rangle \right)^2 \\ & + \alpha \sum_{\substack{\text{电视} \\ \text{我}=1}}^1 \left(\langle w, x_i \rangle - \langle \bar{w}_s, x_i \rangle \right)^2 + \gamma \left(\|\bar{w}\|_2 + \|\bar{w}_s\|_2 \right), \end{aligned} \quad (1)$$

在哪里 $\langle \cdot, \cdot \rangle$ 是内积， \bar{w} 和 \bar{w}_s 是模型的参数₁和模型₂，分别，

尽管 $\alpha, \gamma > 0$ 是正则化系数。这个过程就像“压缩”

来自模型的有用预测信息₁在 $+$ 塑造₂在 $+s$ 。然后，在阶段₂，在

除了模型₃训练基于 $\{-\}$ ，模型₂训练阶段₁仍可使用。

因此，相位预测₂可以通过组合模型制作₃喂和模型₂美联储

与 $+s$ 部分⁷这样，从阶段中吸取了一些经验₁可以被利用

同相₂通过使用模型₂。

有趣的是，即使 $\{-\} \cap X = \emptyset$ ，有些情况下可以启用阶段₁

学习有助于阶段₂，特别是当特征增量发生早于

特征减少[33]，例如，新传感器的部署略早于旧传感器的电池寿命

耗尽。如图 3(c) 所示，在这种情况下，存在一小组数据

这两组特征都可以帮助构建映射 ψ : $\{-\} \rightarrow +$ 。因此，尽管⁸

同相接收₂具有以下特点 $\{-\}$ 仅限，模型₁学习阶段₁仍然可以被利用

通过喂它 ψ ⁹然后，阶段₂可以通过结合模型进行预测₁和

模型₂训练自 $\{-\}$ ，无论是通过加权选择还是加权组合。它具有

已经证明，加权组合的累积损失与

两个模型之间的最小损失，加权选择的累积损失为

与最优选择的损失相当。

这些模型的训练可以通过在线学习技术来完成，例如

在线梯度下降，因此上述策略可以自然地应用于流

数据。值得注意的是，上述策略可以自然地扩展到更多阶段，

并可以通过结合来自不同特征的多个模型进行预测空间。因此，后期阶段的表现甚至可以通过利用集成来增强学习[87]。

最近，有关于在未预训练时使用特征减少/增加进行学习的研究可预测阶段[34]，以及数据分布变化[78]等，以及诸如基于传感器的活动识别[36]。

4. 改变数据分布

近环境机器学习研究通常假设所有数据，包括训练和测试数据是来自相同分布的独立样本（即，**独立同位素**样本）。不幸的是，这并不总是成立。以森林病为例再次提到第 2 节中的监测，该模型可以在夏季基于传感器建立该季节接收到的信号，但希望它在所有季节都能正常工作。图 4 提供了忽略数据分布变化可能会导致严重性能下降。

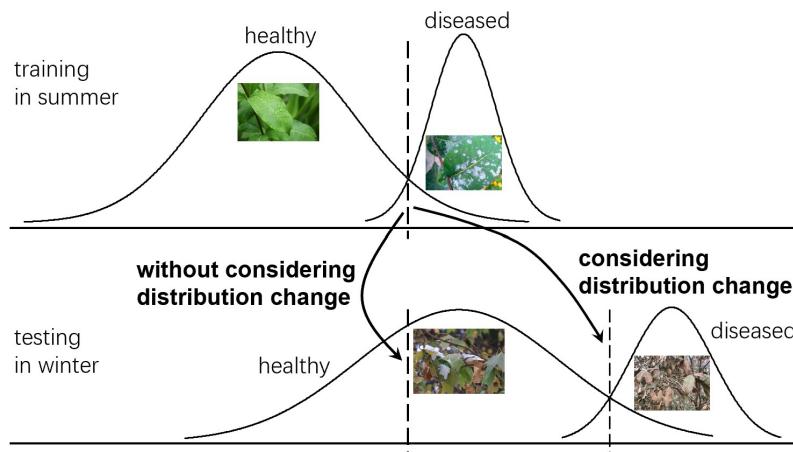


图4：数据分布变化不容忽视。

关于训练/测试数据分布变化的研究已经有很多。例如，先验概率转移和协变量偏移[64]关注磷火车（是/+）=

磷测试（是/十）但磷火车（是） \wedge 磷测试（是）和磷火车（十） \wedge 磷测试（十），而概念漂移[20]担忧磷火车（是/十） \wedge 磷测试（是/十）许多相关研究已经在领域自适应 [37, 4, 40] 或迁移学习 [56] 的指导下进行。注意在蒸汽情况下，数据分布变化可能发生在流的任何阶段，而不是仅限于测试阶段。为了能够处理各种数据分布变化是 Open ML 的一项重要要求。

一般来说，随着数据分布的变化而进行学习并不总是可行的，例如，如果数据分布可以随时任意变化，而无需知道它如何变化变化。幸运的是，在许多实际任务中，可以合理地假设当前观察与最近的观察有密切的关系；换句话说，当前实例和最新的通常来自类似甚至相同的分布，并且远不相似的。因此，我们可以尝试利用流中的一些最新数据来提供帮助。

一般方法通常基于滑动窗口，忘记，或者合奏机制。基于滑动窗口的方法保留最近的实例并丢弃超出范围的旧实例窗口大小固定或自适应 [38, 42]。基于遗忘的方法为每个实例分配权重，并根据实例的年龄降低旧实例的权重 [41, 1]。基于集成 [87] 的方法在集成中添加/删除组件学习器自适应地、动态地调整学习器对于输入实例的权重[28]。

大多数基于滑动窗口或集成的方法需要多次扫描数据。在实际的大数据任务中，往往希望流数据只被扫描一次学习过程所需的存储大小与数据量无关在流结束之前无法知道。最近，一种简单而有效的方法为了解决这个问题，提出了基于遗忘机制的不需要事先了解变化情况，每个实例都可以丢弃一次扫描。此外，受控制理论分析的启发[29]，高概率基于矢量浓度估计误差分析表明，估计误差减少直至收敛。

数据分布变化可能发生在更复杂的情况下，例如丰富的结构。在多示例学习中，有关于这个问题的研究[18]，其中

关键是要考虑包级变化以及实例级变化[77]。

5.多样化的学习目标

学习表现 $f: X \rightarrow Y$ 可以通过绩效衡量来衡量 M_f ，
例如准确率、F1 度量和 ROC 曲线下面积 (AUC)。学习面向不同的
不同的目标可能导致具有不同优势的不同模型。一个模型
在一个指标上最优并不意味着它在其他指标上也能最优。
近环境机器学习研究通常假设 M_f 这将是
用于衡量学习表现的变量应该是不变的并且是预先知道的。
幸运的是，这种情况并不总是成立。以传感器调度任务为例，最初
为了追求高精度的监测，需要部署许多传感器，而
已经达到了相对较高的精度，需要调度其他传感器来确保
系统继续工作，能耗尽可能低。为了能够
处理 Open ML 所需的各种目标。

很少有人研究过具有多样化学习目标的学习。在这里，巨大的挑战
使训练有素的机器学习模型能够顺利地从一个目标切换到另一个目标
另一种方法，无需重新回忆数据来训练一个全新的模型。有研究
基于对许多表演者的观察，将训练好的模型调整到新的目标上
绩效衡量指标是相关的[10, 70]；事实上，各种各样的绩效衡量指标都可以
通过利用非线性辅助分类器进行优化，同时保持较高的计算性能
效率[43]。这也与 模型重用[79, 69]。

除了从一个目标切换到另一个目标之外，学习不同的学习目标
目标也可以通过同时追求多个目标来实现，如果这些目标
目标是事先明确知道的。这需要 帕累托优化来说，
目标是优化最小值 (M_1, M_2, \dots, M_n) 在哪里 M_i 是目标；
更好。通常不存在一个在所有目标上都是最优的模型；
相反，目标是寻求 帕累托前沿 包含绝不逊于其他解决方案的解决方案
解决方案可以同时解决所有目标。图 5 提供了一个示例，其中解决方案

化 + 和是在两个目标上同时不劣于任何其他解决方案，因此它们位于帕累托前沿。

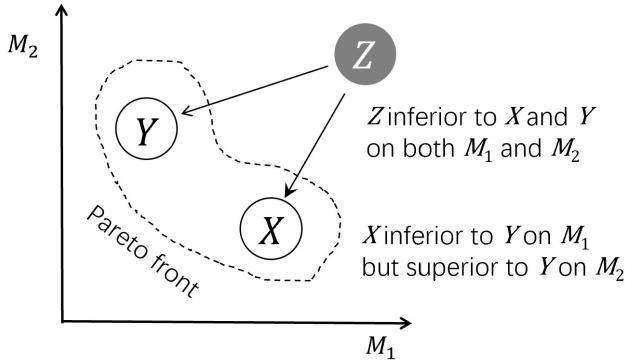


图 5：帕累托前沿的图示。

进化算法，例如遗传算法，通常用于帕累托优化在实践中，尽管它们经常受到批评，因为它们看起来是纯粹的启发式方法，值得一提的是，最近有人试图建立理论进化学习的基础[91]，即利用进化机制，而且已经证明理论进步可以帮助指导强大的新算法的设计，例如进化算法，可以证明比传统算法实现更好的近似保证第一次。

除了显式多目标之外，隐式多目标还需要注意开放机器学习。例如，有些情况下用户无法表达自己的目标显然，但可以提供偏好反馈，如“模型1对我来说比模型2更好”。它已经有证明[16]可以得到这种隐式目标的有效模型通过利用以下技术词袋[75]，假设每个隐含目标都是本质上是一种元素目标的组合。

6.理论问题

Open ML 是一个新的研究方向，因此有太多理论问题需要解决探索。

在图 1 所示的四条线索中，当前新兴的学习技术新类别大多基于启发式方法 [12, 52, 53, 94]。注意，当所有数据是目前有一些理论结果，例如，当新类别存在于未标记数据中时 [76, 46]；然而，当数据积累时，这些结果并不直接适用时间，其中 NewClass 新兴在流中。关于提出了使用递减/增量特征进行学习的算法 [35, 33, 34]，但是缺乏彻底的理论研究。使用进化算法进行多目标学习机制的理论基础已经确立 [91]，但多样化的学习客观问题作为一个整体尚未得到充分探索。随着分布的变化而学习理论研究相对较多。例如，概念漂移的研究思路很长，理论探索 [32, 11, 49]，并提出了一些理论算法分析，从错误和损失界限 [39]、稳定性分析 [30]、泛化和后悔分析 [79] 等。也有关于放宽独立同位素假设 [50, 58, 21]。

Open ML 的挑战性主要在于我们几乎不知道会发生什么变化，以及变化将提前发生。这与强化学习 [65, 47]，学习者与环境互动，探索问题空间。一旦 Open ML 中关注的变化发生，之前的探索就会由于问题空间的改变，强化学习器的知识可能变得无效变化。有研究关于让强化学习器适应变化的环境 [72, 9]，但变化不应该频繁或持续发生。

从技术上讲，在 Open ML 中，没有反映初始训练集，而接收一些实例后必须进行足够的模型更新尽快发生变更。从这个方面来看，Open ML 在某种程度上是相关的到弱监督学习 [89]。然而，与近距离环境研究相比，

强调多数例子，因此通常假设正态分布，在开放ML少数民族例子，甚至是那些从未被观察到的例子很重要，但同时仍然要求大多数人表现出色。因此，与正态分布相比，考虑重尾分布（尤其厚尾分布非常罕见的事件可能会导致极其损失巨大，而不是常见的长尾分布）尾巴不是指数有界，如图6所示。

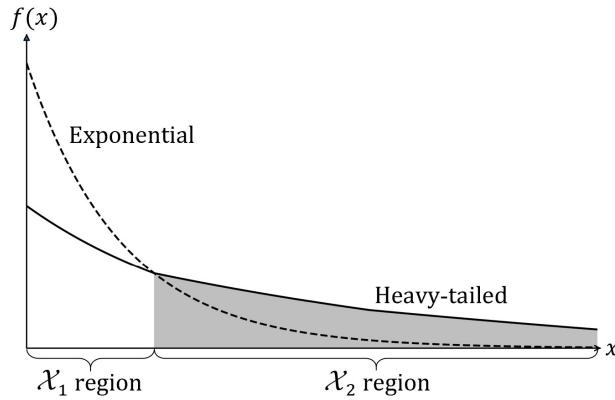


图 6：重尾分布的图示。

显然，人们希望学习到的模型时长 (+) 满足

$$\Pr(\text{Error}(\text{Time}) \leq \varepsilon_{\text{out}}) \geq 1 - \delta_{\text{out}}, \quad (2)$$

在哪里 $\text{Error}(\text{Time}) = \Pr_{x \in \mathcal{X}_{\text{out}}} (\text{Time}_{\text{out}}(+) \neq \text{y}_{\text{out}})$, $i \in \{1, 2\}$, y_{out} 是真实输出 $+$, $0 < \varepsilon_1 \leq \varepsilon_2 \leq \varepsilon$, $\delta_1, \delta_2 < 1$. 直观地讲，这解释了期望的模型应该实现优异的表现 +1 图 6 中的区域（即误差应小于 ε_1 高概率），并令人满意的表现 +2 区域（即错误 ε_2 不得大于比 ε 尽管它可能大于 ε_1 ）严格的门槛 ε 是为了确保最坏的情况无论发生什么变化，性能对于用户来说都是可以承受的。这与安全学习[44]在弱监督场景中，原理优化最坏情况取得良好平均成绩后的表现可以有所帮助。因此，总误差为

$$\text{Error}(\text{Time}) = \varepsilon_1 + \varepsilon_2, \quad (3)$$

在哪里 γ 是权衡的系数 γ_1 和 γ_2 区域，并可由用户根据这些区域的相对重要性； ϵ 根据公式 2，上述理解提供了一种视角来看待 γ_2 区域作为正规化力量学习 γ_1 地区。

典型的重尾分布包括帕累托分布、柯西分布等。当它们被用来代替常用的正态分布时，新的挑战出现。例如，中心极限定理不成立，频繁样本统计量常用的样本均值和方差等统计方法会产生误导（即，它们可能与总体平均值和方差有很大不同）。必须考虑这些问题在 Open ML 中。例如，如果输入和输出空间是重尾的，经验风险最小化变得无效，因为经验风险不再是风险[6]。这对学习算法来说是一个问题，即使对于简单的 L1 回归也是如此[74]。

考虑到随时间积累的数据，需要注意性能指标。这里，令人担忧的是，无论发生什么变化，学习过程仍然在线学习 [7, 62]。与假设静止在线的近距离环境研究相比，线设置，Open ML 关注非平稳在线设置。因此，而不是通过累积损失来衡量表现的静态遗憾学习者与事后选择的最佳常数点的比较，一般动态遗憾[95] 将学习器的累积损失与任意序列的比较结果进行比较，更合理。最近在线凸函数上已经报道了最优结果采用各种机制的优化[73, 84, 82] 和老虎机凸优化[80]。温和条件下非平稳线性老虎机的近似极小最大最优解据报道，通过一个简单而有效的重启机制[83] 方案[68]，它比滑动更适合资源受限的学习任务窗口或遗忘机制。

Open ML 也与噪声数据的学习有关，这方面有很多理论研究，例如 [2, 5, 54, 22, 23]。请注意，与近距离环境研究相比，噪音可以简单地通过平滑等技术来抑制，开放中的重要信号 ML 可能隐藏在被视为噪音的信号中，而罕见的重要事件可能

因过于简单的平滑而感到沮丧。

7. 结论

本文简要介绍了开放环境机器学习的一些研究进展。

它很难对所有相关工作进行彻底的回顾，而且主要是一个简短的总结

作者和他的同事在这个方向上进行了探索，强调一般原则

规则和策略，而不是具体的学习算法。许多策略和想法

本文提到的内容可以通过各种学习技术来实现，可能包括

未来有待探索的不同优势。请注意，讨论的各种问题

在本文中，我们分别讨论这两个问题，但在实际操作中，它们经常同时发生。

让机器学习模型在心理上取得优异表现非常重要

正常情况下，无论发生什么意外不幸，都能保持令人满意的表现

问题。这对于实现强大的人工智能至关重要[14, 15]，并且带来了

学件所需的属性[88]

参考

[1] C. Anagnostopoulos、DK Tasoulis、NM Adams、NG Pavlidis 和 DJ Hand。

具有自适应遗忘功能的在线线性和二次判别分析

分类。*统计分析与数据挖掘*, 5(2): 139–166, 2012.

[2] D. Angluin 和 P. Laird. 从嘈杂的例子中学习。*机器学习*, 2(4):343–

370, 1988年。

[3] PL Bartlett 和 MH Wegkamp。使用铰链的拒绝选项分类

损失。*机器学习研究杂志*, 9(59): 1823–1840, 2008.

[4] S. Ben-David, T. Lu, T. Luu, 和 D. Pal. 领域自适应的不可能性定理

化。在*美国统计局*, 第 129–136 页, 2010 年。

[5] A. Blum, A. Kalai, 和 H. Wasserman. 噪声容忍学习、奇偶校验问题、

以及统计查询模型。*ACM 杂志*, 50(4):506–519, 2003.

- [6] O. Catoni. 挑战经验均值和经验方差: 一项偏差研究。
亨利·庞加莱概率与统计研究所年鉴, 48(4):1148–1185,
2012 年。
- [7] N. Cesa-Bianchi 和 G. Lugosi。预测、学习和游戏. 剑桥大学
城市出版社, 2006。
- [8] J. Chen, Y. Geng, Z. Chen, I. Horrocks, JZ Pan 和 H. Chen. 知识感知
零样本学习: 综述与展望。在国际计算机辅助教学, 第 4366–4373 页, 2021 年。
- [9] S.-Y. Chen, Y. Yu, Q. Da, J. Tan, H.-K. Huang, 和 H.-H. Tang. 稳定加固
动态环境中的学习及其在在线推荐中的应用。在
知识发现, 第 1187–1196 页, 2018 年。
- [10] C. Cortes 和 M. Mohri。AUC 优化与错误率最小化。在神经信息处理系统,
第 313–320 页, 2004 年。
- [11] K. Crammer, Y. Mansour, E. Even-Dar 和 JW Vaughan。遗憾最小化
概念漂移。在小马, 第 168–180 页, 2010 年。
- [12] Q. Da, Y. Yu, 和 Z.-H. Zhou. 利用未标记的样本进行增强类学习
数据。在美国航空学会联合会, 第 1760–1766 页, 2014 年。
- [13] M. Delange、R. Aljundi、M. Masana、S. Parisot、X. Jia、A. Leonardis、G. Slabaugh,
和 T. Tuytelaars。持续学习调查: 在课堂上抵制遗忘
化任务。IEEE 模式分析与机器智能, 页
10.1109/TPAMI.2021.3057446, 2022。
- [14] TG Dietterich.迈向强健人工智能的步骤。人工智能杂志, 38(3):3–24,
2017 年。
- [15] TG Dietterich.健壮的人工智能和健壮的人类组织。正面
计算机科学层次, 13(1):1–3, 2019。
- [16] Y.-X. Ding 和 Z.-H. Zhou. 基于偏好的学习目标自适应。
神经成像与成像系统, 第 7839–7848 页, 2018 年。

- [17] ER Faria、IR Gonçalves、J. Gama 和 ACPLF Carvalho。评价
数据流的多类新颖性检测算法。*IEEE 知识与
数据工程*《临床神经病学杂志》2015年第27卷第11期
- [18] J. Foulds 和 E. Frank. 多示例学习假设的回顾。*知识
工程评论*, 25(1):1-25, 2010。
- [19] G. Fumera、F. Roli 和 G. Giacinto. 具有多个阈值的拒绝选项。*图案
识别*, 33(12): 2099–2101, 2000年。
- [20] J. Gama、I. Zliobaite、A. Bifet、M. Pechenizkiy 和 A. Bouchachia。
概念漂移适应。*ACM 计算调查*《中国药物滥用状况评估报告》2014年第44期
- [21] W. Gao, X.-Y. Niu, 和 Z.-H. Zhou. 非独立同分布的可学习性急性髓系白血病, 第 158 页–
173, 2016年。
- [22] W. Gao, L. Wang, Y.-F. Li, 和 Z.-H. Zhou. 风险最小化存在
标签噪音。在*美国航空学会联合会*, 第 1575–1581 页, 2016 年。
- [23] W. Gao, T. Zhang, B.-B. Yang, and Z.-H. Zhou. 关于噪声估计的统计。
人工智能, 293: 文章103451, 2021年。
- [24] Y. Geifman 和 R. El-Yaniv. SelectiveNet: 一种集成
拒绝选项。在*国际激光医学联合会*, 第 2151–2159 页, 2019 年。
- [25] C. Geng, S.-J. Huang, 和 S. Chen. 开放集识别的最新进展: 综述。
IEEE 模式分析与机器智能《临床神经病学杂志》2019年第43卷, 第10期, 3614–3631页。
- [26] C. Giraud-Carrier. 关于增量学习效用的注释。*人工智能通讯*,
13(4):215–223, 2000年。
- [27] T. Golub、D. Slonim、P. Tamayo、C. Huard、M. Gaasenbeek、J. Mesirov、H. Coller,
M. Loh、J. Downing、M. Caligiuri、C. Bloomfield 和 E. Lander。分子分类
癌症的确定：通过基因表达进行类别发现和类别预测。*科学*,
286(5439):531–537, 1999年。

- [28] HM Gomes, JP Barddal, F. Enembreck 和 A. Bifet. 集成研究
数据流分类的学习。*ACM 计算调查*, 50(2):23:1–23:36,
2017年。
- [29] L. Guo, L. Ljung, 和 P. Priouret. 遗忘因子 RLS 的性能分析
算法。*国际自适应控制信号处理杂志*, 7(6):525–538,
1993年。
- [30] M. Harel, S. Mannor, R. El-Yaniv, 和 K. Crammer. 通过
重采样。在*国际激光医学联合会*, 第 1009–1017 页, 2014 年。
- [31] H. He, S. Chen, K. Li, 和 X. Xu. 基于流数据的增量学习。电气电子工程师学会
跨神经网络《自然》2011年第22期, 22(12): 1901–1914.
- [32] DP Helmbold 和 PM Long。通过最小化分歧来跟踪漂移概念
的评论。*机器学习*, 14(1):27–45, 1994.
- [33] B.-J. Hou、L. Zhang 和 Z.-H. Zhou。使用特征可演化流进行学习。在
神经信息处理系统, 第 1417–1427 页, 2017 年。
- [34] B.-J. Hou, L. Zhang, 和 Z.-H. Zhou. 具有不可预测特征的预测
真正的进化。 *IEEE 跨神经网络和学习系统*, 页
DOI: 10.1109/TNNLS.2021.3071311, 2021。
- [35] C. Hou 和 Z.-H. Zhou. 具有增量和减量特征的一次性学习。
IEEE 模式分析与机器智能, 40(11): 2776–2792, 2018。
- [36] C. Hu, Y. Chen, X. Peng, H. Yu, C. Gao 和 L. Hu. 新功能增量
基于传感器的活动识别的学习方法。 *IEEE 知识与*
数据工程《细胞与分子生物学杂志》, 31(6): 1038–1050, 2019年。
- [37] H. Daumé III 和 D. Marcu. 统计分类器的领域自适应。*期刊/人工智能研究*, 26: 101–126, 2006。
- [38] R. Klinkenberg 和 T. Joachims. 使用支持向量机检测概念漂移。
在*国际激光医学联合会*, 第 487–494 页, 2000 年。

- [39] JZ Kolter 和 MA Maloof。利用加性专家集成来应对概念漂移。在国际激光医学联合会, 第 449–456 页, 2005 年。
- [40] WM Kouw 和 M. Loog. 无目标标签的领域自适应综述。
IEEE 模式分析与机器智能, 43(3):766–785, 2005。
- [41] I. Koychev. 逐渐遗忘以适应概念漂移。*ECAI-2000 工作时空推理中的当前问题*, 第 101–106 页, 2000 年。
- [42] LI Kuncheva 和 I. Zliobaite. 关于窗口大小在分类过程中的变化环境。*智能数据分析*, 13(6):861–872, 2009.
- [43] N. Li, IW Tsang, 和 Z.-H. Zhou. 绩效指标的有效优化通过分类器自适应。*IEEE 模式分析与机器智能*, 35(6):1370–1382, 2013.
- [44] Y.-F. Li 和 Z.-H. Zhou. 致力于让未标记数据永远不会受到伤害。*IEEE 译文模式分析与机器智能*, 37(1): 175–188, 2015。
- [45] FT Liu, KM Ting, 和 Z.-H. Zhou. 隔离森林。*信息通信技术*, 第 413–422 页, 2008 年。
- [46] S. Liu、R. Garrepalli、D. Hendrycks、A. Fern、D. Mondal 和 TG Dietterich。PAC 检测新类别的保证和有效算法。*机械工程学报学习研究*, 23(44):1–47, 2022。
- [47] AY Majid、S. Saaybi、T. van Rietbergen、V. Francois-Lavet、RV Prasad 和 C. Verhoeven。深度强化学习与进化策略：比较调查。*arXiv:2110.01411*, 2021 年。
- [48] M. Masana、X. Liu、B. Twardowski、M. Menta、AD Bagdanov 和 J. van de Weijer。类增量学习：图像分类的调查与性能评估。
arXiv:2010.15277, 2021 年。
- [49] M. Mohri 和 AM Medina. 漂移学习的新分析和算法分布。在*谷氨酰胺*, 第 124–138 页, 2012 年。

- [50] M. Mohri 和 A. Rostamizadeh。非独立同分布函数的 Rademacher 复杂度界限过程。在神经信息处理系统, 第 1097–1104 页, 2008 年。
- [51] S. Monti、P. Tamayo、J. Mesirov 和 T. Golub。共识聚类：重采样-基于基因表达微阵列数据的类别发现和可视化方法。*机器学习*, 52(1-2):91–118, 2003.
- [52] X. Mu, KM Ting, 和 Z.-H. Zhou. 流媒体下的新兴技术分类类别：使用完全随机树的解决方案。*IEEE 知识与数据汇刊工程*《细胞与分子生物学杂志》, 29(8): 1605–1618, 2017.
- [53] X. Mu, F. Zhu, J. Du, E.-P. Lim, 和 Z.-H. Zhou. 使用流式分类新兴类别的矩阵草图。在美国航空学会联合会, 第 2373–2379 页, 2017 年。
- [54] N. Natarajan, IS Dhillon, PK Ravikumar 和 A. Tewari. 学习噪音标签。在神经信息处理系统, 第 1196–1204 页, 2013 年。
- [55] S. Ozawa, SL Toh, S. Abe, S. Pang, 和 N. Kasabov. 特征增量学习用于人脸识别的空间和分类器。*神经网络*《细胞与分子生物学杂志》2008年第19期, 18(5-6):575–584, 2005.
- [56] SJ Pan 和 Q. Yang. 迁移学习调查。*IEEE 知识与数据工程*《细胞与分子生物学杂志》2009年第22卷第10期
- [57] J. Parmar、SS Chouhan、V. Raychoudhury 和 SS Rathore。开放世界机器学习：应用、挑战和机遇。arXiv:2105.13448, 2021 年。
- [58] A. Pentina 和 CH Lampert。非独立同分布任务的终身学习。*神经信息处理系统*、页 1540–1548, 2015年。
- [59] B. Pfürlb 和 A. Gepperth。灾难性灾害的综合应用研究 DNN 中的遗忘。在国际肾病研究联合会, 2019年。
- [60] WJ Scheirer、AR Rocha、A. Sapkota 和 TE Boult。面向开放集识别化。*IEEE 模式分析与机器智能*《细胞与分子生物学杂志》, 2013, 35(7): 1757–1772。
- [61] V. Sehwag、AN Bhagoji、L. Song、C. Sitawarin、D. Cullina 和 M. Chiang。分析开放世界机器学习的稳健性。在第 12 届 ACM 人工智能研讨会情报与安全, 第 105–116 页, 2019 年。

- [62] S. Shalev-Shwartz. 在线学习和在线凸优化。基金会和机器学习的趋势, 4(2): 107–194, 2011。
- [63] R. Socher、M. Ganjoo、CD Manning 和 AY Ng。通过跨模式换乘。神经信息处理系统, 第 935–943 页, 2013 年。
- [64] M. Sugiyama 和 M. Kawanabe。非平稳环境中的机器学习: 协变量移位适应简介. 麻省理工学院出版社, 马萨诸塞州剑桥, 2012 年。
- [65] RS Sutton 和 AG Barto。强化学习: 简介. 麻省理工学院出版社, 马萨诸塞州剑桥, 第二版, 2012 年。
- [66] N. Syed、H. Liu 和 K. Sung。使用支持向量机进行增量学习。IJCAI-99 支持向量机研讨会, 1999 年。
- [67] PE Utgoff. 决策树的增量归纳。机器学习, 4(2): 161–186, 1989 年。
- [68] C.-Y. Wei 和 H. Luo. 无需先验知识的非平稳强化学习: 一种最优黑盒方法。小马, 第 4300–4354 页, 2021 年。
- [69] X.-Z. Wu, S. Liu, 和 Z.-H. Zhou. 通过优化多党派多阶级边缘。国际激光医学联合会, 第 6840–6849 页, 2019 年。
- [70] X.-Z. Wu 和 Z.-H. Zhou. 多标签绩效衡量的统一视图。国际激光医学联合会, 第 3780–3788 页, 2017 年。
- [71] Y. Xian, CH Lampert, B. Schiele 和 Z. Akata. 零样本学习 - 综合评价好、坏、丑。IEEE 模式分析与机器智能《细胞与分子生物学杂志》, 41(9): 2251–2265, 2019 年。
- [72] C. Zhang, Y. Yu, 和 Z.-H. Zhou. 学习环境校准动作政策自我演进。在国际计算机辅助教学, 第 3061–3067 页, 2018 年。
- [73] L. Zhang, S. Lu, 和 Z.-H. Zhou. 动态环境中的自适应在线学习。在神经成像与成像系统, 第 1330–1340 页, 2018 年。

- [74] L. Zhang 和 Z.-H. Zhou. 重尾分布的 L1 回归。在 *神经成像与成像系统*, 第 1084–1094 页, 2018年。
- [75] Y. Zhang, R. Jin, 和 Z.-H. Zhou. 理解词袋模型：统计框架。*国际机器学习与控制论杂志*, 1(1):43–52, 2010 年。
- [76] Y.-J. Zhang, P. Zhao, L. Ma, 和 Z.-H. Zhou. 学习的无偏风险估计器具有增强类。在 *神经成像与成像系统*, 第 10247–10258 页, 2020 年。
- [77] Y.-L. Zhang 和 Z.-H. Zhou. 带关键实例转移的多示例学习。*国际计算机辅助教学*, 第 3441–3447 页, 2017年。
- [78] Z.-Y. Zhang, P. Zhao, Y. Jiang, 和 Z.-H. Zhou. 基于特征和分布的学习可演化流。在 *国际激光医学联合会*, 第 11317–11327 页, 2020 年。
- [79] P. Zhao, L.-W. Cai, 和 Z.-H. Zhou. 通过模型重用处理概念漂移。*机器学习*, 109(3): 533–568, 2020年。
- [80] P. Zhao, G. Wang, L. Zhang, 和 Z.-H. Zhou. 非函数中的 Bandit 凸优化固定环境。*机器学习研究杂志*, 22(125):1–45, 2021。
- [81] P. Zhao, X. Wang, S. Xie, L. Guo, 和 Z.-H. Zhou. 无分布一次性学习。*IEEE 知识与数据工程汇刊*, 33(3): 951–963, 2021年。
- [82] P. Zhao, Y.-X. Wang, 和 Z.-H. Zhou. 具有记忆的非平稳在线学习和非随机控制。在 *美国统计局*, 第 2101–2133 页, 2022 年。
- [83] P. Zhao, L. Zhang, Y. Jiang, 和 Z.-H. Zhou. 一种简单的非平稳方法线性匪徒。在 *美国统计局*, 第 746–755 页, 2020 年。
- [84] P. Zhao, Y.-J. Zhang, L. Zhang, 和 Z.-H. Zhou. 凸和光滑的动态遗憾函数。在 *神经成像与成像系统*, 第 12510–12520 页, 2020 年。
- [85] P. Zhao, Y.-J. Zhang, 和 Z.-H. Zhou. 具有未知未知数。在 *美国航空学会联合会*, 第 10999–11006 页, 2021 年。

- [86] G. Zhou, K. Sohn, 和 H. Lee. 带去噪的在线增量特征学习自动编码器。在美国统计局, 第 1453–1461 页, 2012 年。
- [87] 周志华。集成方法：基础和算法. 查普曼与霍尔/CRC, 佛罗里达州博卡拉顿, 2012 年。
- [88] Z.-H. Zhou. Learnware: 论机器学习的未来。计算机前沿科学, 10(4):589–590, 2016。
- [89] Z.-H. Zhou. 弱监督学习简介。国家科学审查, 5(1):44–53, 2018。
- [90] Z.-H. Zhou 和 Z. Chen. 混合决策树。知识系统, 15(8):515–528, 2002年。
- [91] Z.-H. Zhou, Y. Yu, 和 C. Qian. 进化学习：理论与实践进展算法. Springer, 柏林, 2019 年。
- [92] Y. Zhu, KM Ting, 和 Z.-H. Zhou. 在多实例中发现多个新标签多标签学习。在美国航空学会联合会, 第 2977–2983 页, 2017年。
- [93] Y. Zhu, KM Ting, 和 Z.-H. Zhou. 通过实例生成实现新类自适应一次性增量学习。在信息通信技术, 第 1207–1212 页, 2017年。
- [94] Y. Zhu, KM Ting, 和 Z.-H. Zhou. 新兴标签的多标签学习。IEEE 知识与数据工程汇刊/《自然》杂志, 30(10): 1901–1914, 2018.
- [95] M. Zinkevich. 在线凸规划和广义无穷小梯度分国际激光医学联合会, 第 928–936 页, 2003 年。