**Generalizing from a Few Examples: A Survey on Few-Shot Learning**

人工智能在数据密集型应用中取得了成功，但它缺乏从有限的例子中学习的能力。为了解决这一问题，提出了少镜头学习（FSL）。利用先验知识，可以快速地从有限监督经验的新任务中归纳出来。为了全面了解FSL，我们进行了一项调查研究。我们首先澄清了FSL的正式定义。然后指出不可靠经验风险最小化是FSL的核心问题。基于如何利用先验知识来处理核心问题，我们将不同的FSL方法分为三类：数据利用先验知识来增加监督经验，模型利用先验知识来约束假设空间，算法利用先验知识来改变对最优假设参数的搜索在假设空间里。在这种统一的分类法下，我们对不同类别的利弊进行了深入的讨论。最后，从问题设置、技术、应用和理论等方面提出了可行性研究方向，希望能为后续研究提供一些启示。

附加关键词和短语：少镜头学习、单镜头学习、低镜头学习、小样本学习、元学习、先验知识

**1引言**

“机器能思考吗[122]？这是1950年艾伦·图灵（Alan Turing）题为“计算机械与智能”的开创性论文中提出的问题。他说，“数字计算机背后的想法可以解释为，这些机器旨在执行任何可以由人类计算机完成的操作”。换句话说，机器的最终目标是和人类一样聪明。近年来，由于GPU等强大计算设备的出现，ImageNet等大规模数据集[26]，CNN等先进模型和算法[64]，人工智能加快了向人类靠拢的步伐，在许多领域击败了人类。举几个例子，AlphaGo[106]在玩古老的围棋游戏中击败了人类冠军，ResNet[50]在1000个类的ImageNet数据集上的分类精度高于人类。而在其他领域，人工智能作为高度智能的工具，如语音助手、搜索引擎、自动驾驶汽车和工业机器人，参与到人类的日常生活中。

尽管人工智能很繁荣，但目前的人工智能还不能从几个例子中迅速概括出执行任务的方法。上述人工智能的成功应用依赖于对大规模数据的详尽学习。相比之下，人类能够通过利用过去所学的知识快速地学习新的任务场景。例如，一个学会了加法的孩子，只要举几个例子，就可以迅速地把知识转移到乘法上，例如2×3=2+2+2和1×3=1+1+1。另一个例子是，给一个陌生人的几张照片，一个孩子可以很容易地从大量的照片中辨认出同一个人。

弥合人工智能和类人学习之间的鸿沟是一个重要的方向。这可以通过机器学习来解决，机器学习是人工智能的一个分支领域，它支持诸如模型、算法和理论等科学研究基础。具体来说，机器学习关注的是如何构造能够随着经验自动改进的计算机程序的问题[80]。

为了满足从有限的有监督信息中学习以获得任务挂起的渴望，提出了一个新的机器学习问题，称为少镜头学习（FSL）[32，33]。当只有一个例子需要学习时，FSL也被称为一次性学习问题。通过融合先验知识，FSL可以学习有限监督信息的新任务。

FSL是人工智能的试验台。因此，它是否像人类一样学习是一个考验。一个典型的例子是字符识别[66]，在这里计算机程序被要求分类，解析和生成新的手写字符给出了几个例子。要处理这个任务，可以将字符分解成可以跨字符转移的较小部分，然后将这些较小的组件聚合成新字符。这是一种像人一样的学习方式[67]。自然，FSL推动了机器人学的发展[24]，其目标是开发能够复制人类行为的机器，以便在某些情况下取代人类。例如单次模拟[30]、多武装匪徒[30]、视觉导航[30]、移动中的连续控制[34]。

除了测试人工智能之外，FSL还可以帮助减轻为工业需要收集大规模监督数据的负担。例如，ResNet[50]在1000个类的ImageNet数据上获得了比人类更高的分类精度。但是，这是在每个类都有足够的标记图像的情况下发生的。相比之下，人类可以识别大约30000个类[16]，其中收集每个类的足够图像供机器学习是非常困难的。

相反，FSL可以帮助减少这些数据密集型应用的数据收集工作，例如图像分类[127]、图像检索[118]、对象跟踪[15]、手势识别[88]、图像字幕和视觉问答[28]、视频事件检测[137]和语言建模[127]。此外，能够执行FSL可以降低那些计算成本高昂的应用程序的成本，例如一次性架构搜索[21]。当模型和算法成功地用于FSL时，它们自然地适用于许多样本的数据集，这些样本集更容易学习。

FSL的另一个经典场景是，由于隐私、安全或伦理问题等原因，很难或不可能获取受监督信息的任务。例如，药物发现就是发现新分子的特性，从而将有用的分子识别为新药的过程[3]。然而，由于可能的毒性、低活性和低溶解度，这些新分子在临床候选者中没有很多真实的生物学记录。这使得药物发现任务成为一个FSL问题。类似的罕见案例学习应用程序可以是FSL翻译[56]，冷启动项目建议[126]，其中目标任务没有很多示例。正是通过FSL，学习适合这些罕见情况的模型成为可能。

随着人工智能的学术梦想和廉价学习的工业需求，FSL引起了人们的广泛关注，成为一个热门话题。作为一种学习范式，许多方法都试图解决这一问题，如元学习方法[100]、嵌入学习方法[127]和生成建模方法[31]。然而，目前还没有工作提供一个有组织的分类法来连接FSL方法，解释为什么有些方法有效而另一些方法失败，也没有讨论不同工作的利弊。因此，我们对FSL问题进行了调查。本次调查的贡献总结如下。

•我们给出了FSL的正式定义。它可以自然地链接到[80]中提出的经典机器学习定义。这个定义不仅足够笼统，足以包括所有现有的FSL工作，但也足以明确什么是FSL的目标，以及我们如何解决它。这样的定义有助于确定未来FSL领域的研究目标。

指出了机器学习中基于误差分解的FSL的核心问题。我们发现，正是不可靠的经验风险最小化使得FSL难以学习。这可以通过满足或降低学习的样本复杂度来缓解。更重要的是，这为以更有组织和系统的方式改进FSL方法提供了见解。

•我们对从FSL诞生到最近出版的文献进行了广泛的文献回顾，并根据数据、模型和算法对它们进行了统一分类。对不同类别的利弊进行了深入的讨论。我们还总结了每个类别下的见解。这有助于更好地理解FSL方法。

•我们从问题设置、技术、应用和理论方面为FSL提出了四个有前途的未来方向。这些见解是基于目前FSL发展的弱点，以及未来可能的改进。我们希望他们能提供一些见解。

**1.1调查的组织**

本次调查的其余部分组织如下。第二节概述了本研究的概况，包括FSL的形式定义、核心问题、相关的学习问题以及现有研究在数据、模型和算法方面的分类。第3节是关于增加数据以解决FSL问题的方法。第4节是用于约束模型以使FSL可行的方法。第五节是改变算法搜索策略来处理FSL问题的方法。在第六部分中，我们从问题设置、技术、应用和理论等方面提出了FSL的未来发展方向。

最后，调查在第7节得出结论。

**2概述**

在本节中，我们首先在第2.1节中提供本文中使用的符号。第2.2节给出了FSL问题的形式化定义，并给出了具体的例子。由于FSL问题涉及到许多机器学习问题，我们在第2.3节讨论了它们之间的联系和区别。在第2.4节中，我们揭示了使FSL问题变得困难的核心问题。然后根据现有作品如何处理核心问题，我们在第2.5节中提出了一个统一的分类法。

**2.1表示法**

考虑一个有监督的学习任务T，FSL处理一个数据集D={}，该数据集由训练集组成，其中i很小，而测试集。通常，人们会考虑N-way-K-shot分类任务[34127]，其中包含来自N个类的I=KN示例，每个类都有K个示例。设p（x，y）为输入x和输出y的基真联合概率分布，为x到y的最优假设，FSL通过拟合和测试来学习发现。为了逼近，模型确定由参数化的假设h（·；θ）的假设空间。算法是通过搜索来寻找参数化最优h∈的θ的优化策略。性能是通过在预测（例如=h（x；θ））和实际输出y上定义的损失函数（yˆ，y）来测量的。

**2.2问题定义**

由于FSL自然是机器学习的一个子领域，在给出FSL的定义之前，让我们回顾一下机器学习是如何被定义的，这里我们采用Mitchell的定义[80]，如定义2.1所示。

定义2.1（机器学习[80]）。如果一个计算机程序的性能能随着用P度量的E-on-T而提高，则它可以从一些任务T和性能度量P的经验E中学习。

如我们所见，一个机器学习问题是由E，T和P指定的。例如，考虑到图像分类任务（T），机器学习程序可以通过使用大规模标记图像（例如ImageNet data set）训练获得的E来提高其分类精度（P）【64】。

另一个例子是最近的计算机程序AlphaGo[106]，它在玩古老的围棋（T）游戏时击败了人类冠军。它通过使用一个记录了3000万次人类专家动作的数据库进行训练，以及反复与自己比赛，提高了对对手的胜率（P）。

上述典型的机器学习应用对给定的任务需要大量的监督信息。然而，如导言所述，这可能是困难的，甚至是不可能的。FSL是机器学习的一个特例，它的目标是在数据集D提供有限的监督信息的情况下获得良好的学习性能。监督信息是指训练数据集包括输入及其相应的输出的示例[17]。正式地，我们在定义2.2中定义了FSL。

**定义2.2**。少镜头学习（FSL）是一类机器学习问题（由E、T和P指定），其中E包含目标T的少量监督信息。

为了更好地理解这个定义，让我们展示FSL的三个典型场景（表1）：

•充当类似人类学习的试验台：要走向人类智能，计算机程序解决FSL问题的能力至关重要。一个流行的任务（T）是生成一个新字符的样本，只给出几个例子[66]。受人类学习方式的启发，计算机程序使用E进行学习，E既包括作为监督信息的给定示例，也包括作为先验知识的预先训练的概念，例如部分和关系。通过视觉图灵测试（P）的通过率对生成的字符进行评价，判断图像是由人生成的还是由机器生成的。有了这些先验知识，计算机程序还可以学习分类、解析和生成一些像人类这样的例子的新手写字符。

•减少数据收集工作和计算成本：FSL还有助于减轻收集大规模监督信息的负担。考虑通过FSL[32]对几个示例的类进行分类。对于目标T的每一类，用少量标记图像得到的E，以及从其他类中提取的先验知识（如原始图像）进行协同训练，提高了图像分类精度（P）。成功完成此任务的方法通常具有较高的通用性，因此它们可以很容易地应用于多个样本的任务。

•针对罕见情况的学习：最后，通过FSL，人们可以学习适用于受监督数据有限的罕见情况的模型。例如，考虑一个常见的药物发现任务（T），即预测新分子是否会产生毒性作用[3]。正确分配到有毒或非有毒（P）的分子百分比随着新分子有限分析和许多类似分子的先前知识分析获得的E而提高。

由于E中只含有少量与T直接相关的监督信息，因此一般的监督机器学习方法在FSL问题上都是失败的。因此，FSL方法通过将E中可用的监督信息与一些先验知识（即“学习者在看到示例之前关于未知功能的任何信息”）相结合，使目标T的学习成为可行的[75]。

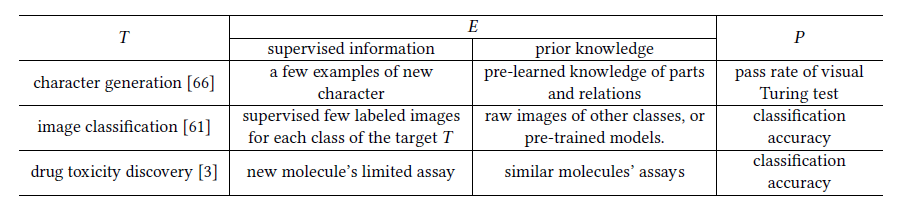


表1。基于定义2.2的三个FSL示例的插图。

**2.3相关学习问题**

在本节中，我们将讨论FSL的相关学习问题。特别阐明了与FSL的联系和区别。

•半监督学习[147]通过由标记样本和未标记样本组成的经验E学习最优假设。正无标记学习[71]是半监督学习的一个特例，它只给出正样本和无标记样本。另一个相关的半监督学习问题是主动学习[103]，它选择信息性的未标记数据来查询oracle的输出y。根据定义，FSL可以是监督学习、半监督学习和强化学习，这取决于除了有限的监督信息之外还有什么样的数据可用。它既不需要存在未标记的样本，也不需要oracle。

•不平衡学习[49]从经验E中学习，y的分布严重倾斜。它训练和测试在所有可能的y中进行选择。相反，FSL训练和测试仅举几个例子，而可能将其他y作为学习的先验知识。

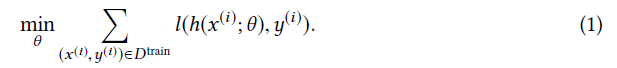
•转移学习[87]将从有足够训练数据的源域和源任务学到的知识转移到训练数据有限的目标域和目标任务。领域适应[11]是一种迁移学习问题，任务相同，领域不同。另一个相关的转移学习问题称为零射击学习[68]，它通过将没有监督的训练示例链接到现有类来识别新类，这些类通常依赖外部数据源，如文本语料库和词汇数据库[135]。FSL不需要是一个迁移学习问题。然而，当给定的监督信息被限制为直接学习时，FSL需要将先验知识转移到当前任务中。然后这种FSL问题就变成了转移学习问题。

•元学习或learning-to-learn [51]通过提供的数据集和元学习者跨任务提取的元知识，提高了新任务T的P。具体来说，元学习者在任务间逐渐学习一般信息（元知识），并且学习者使用任务特定信息快速地将元学习者概括为新任务T。许多FSL方法都是以元学习者为先验知识的元学习方法。为了以后的参考，元学习的正式定义见附录a。

**2.4核心问题**

通常，对于机器学习问题，我们无法得到完美的预测，即存在一些预测误差。在这一部分中，我们说明了机器学习中基于错误分解的FSL下的核心问题[19，20]。

回想一下，机器学习是关于用P来衡量E对T的改进。根据我们的符号，这可以写成



这里，θ参数化了模型选择的假设h∈。学习的内容是寻找符合数据的中最佳假设h的θ的算法。

**2.4.1经验风险最小化**。实质上，我们希望最小化预期风险R，即相对于p（x，y）的损失。对于某些假设h，R被定义为

然而，p（x，y）是未知的。因此，经验风险被用来估计预期风险R（h）。

定义为训练数据集（I个样本的）上样本损失的平均值：

而学习是通过经验风险最小化[124]来完成的（也许还需要一些正则化）。

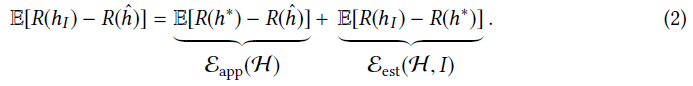
为便于说明，设

•，其中R达到其最小值；

•，其中R相对于h∈最小化；

•，其中RI相对于h∈最小化。

假设、和是唯一的，因为它们很简单。对于随机选择的训练集，学习的总误差可以分解为



当近似误差测量H中的函数接近最优假设的程度时，估计误差测量H中最小化经验风险而不是预期风险R（H）的效果[19，20]。

如上所示，总误差受（假设空间）和I（中的示例数）的影响。换言之，可以尝试从提供的数据、确定的模型和通过搜索最适合的h的θ的算法的角度来学习减少总误差。

**2.4.2不可靠的经验风险最小化**。注意，对于（2）中的，我们有



这意味着更多的例子可以帮助减少。因此，在监督学习任务的公共设置中，训练数据集具有足够的监督信息，即I是大的。

经验风险最小化者可以为中的h提供一个很好的（即，根据（3））近似到尽可能好的。

但是，我在FSL中可用的示例数量很少。这使得经验风险远不能很好地逼近预期风险R（h），由此产生的经验风险最小化也不好。事实上，这是FSL的核心问题，即经验风险最小化不再可靠。因此，FSL比一般的机器学习设置困难得多。图1显示了普通和少量放炮设置之间的比较。

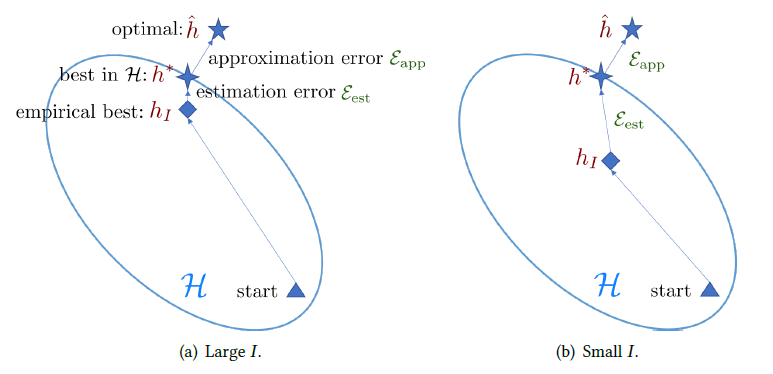


图1。机器学习中常用和少镜头设置的比较。

从历史上看，经典的机器学习方法是通过正则化来学习的[41]，以推广新数据集的学习方法。正则化技术植根于机器学习，有助于减少并获得更好的学习性能[80]。经典的例子包括Tikhonov正则化器[52]和lasso正则化器[116]。不可否认，这些正则化可以限制模型的形式。然而，这些简单的正则化技术并不能解决FSL问题。它们不会带来任何额外的监督信息，因此它们无法解决小数据量导致的经验风险最小化的不可靠性。因此，正则化学习不足以为FSL问题提供良好的预测性能。

**2.4.3样本复杂度**。经验风险最小化与样本复杂度密切相关。具体地说，样本复杂度是指保证损失最小的经验风险而不是预期风险所需的训练样本的数量最多为概率为1-δ[80]。数学上，对于0<ϵ，δ<0.5，样本复杂度是一个整数S，因此对于I≥S，我们有



当S是有限的时，是可学习的。对于无限空间，其复杂度可以用Vapnik–Chervonenkis（VC）维数来度量[125]。VC维VC（）定义为可以被分解（以所有可能的方式拆分）的最大输入集的大小



其中上界和下界分别在[125]和[114]中得到证明。

如（4）和（5）所示，对于固定的δ和ϵ，需要较少的复杂度来使所提供的I样本足够S。FSL方法通常使用先验知识来补偿样本不足。一种典型的FSL方法是贝叶斯学习[32，66]。它将所提供的训练数据集与先验概率（给出之前可用的概率）结合起来[17]。这样，确定h的最终概率所需的S可证明地减小了[39，80]。这启发我们怀疑FSL方法可以通过利用先验知识来满足或减少S。从而解决了不可靠经验风险最小化的核心问题。

**2.5分类**

在前面的章节中，我们发现FSL的核心问题是不可靠的经验风险最小化。我们还证明，通过满足或降低样本复杂度，可以获得可靠的经验风险最小化。

根据如何使用先验知识，我们将这些工作分类为：

•数据：利用先验知识将I样本的增加到样本的方法。这样，S就可以满足[12,107]。如图2（a）所示，可以得到更准确的经验风险最小化。随着样本的增多，常用的模型和算法可以直接使用。

•模型：基于经验E中的先验知识设计以限制的复杂性的方法。使用这种限制的学习导致更小的S，如[39，75，85]所证明的。如图2（b）所示。灰色区域不考虑以后的优化，因为根据先前的知识，它们不太可能包含最优。对于这个较小的，足够学习一个更可靠的。

•算法：利用先验知识搜索中参数化最佳假设的θ的方法。先验知识通过提供良好的开始搜索的初始点或直接提供搜索步骤来改变搜索策略。对于这类方法，改进现有的一些参数可以优化和目标，同时利用从一组任务中学习到的元学习者直接针对目标，因此我们在图2（c）中显示了从“开始”到的两条路径。它们都可以证明S[476]。

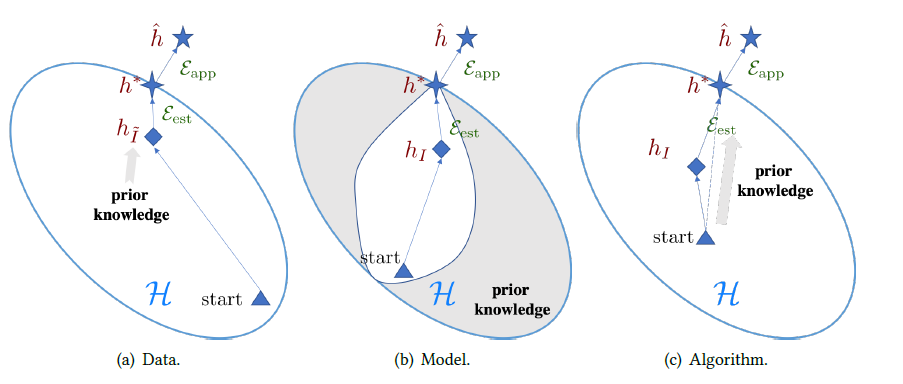


图2。FSL方法如何从数据（左）、模型（中）和算法（右）三个方面解决少镜头问题。在图2（a）中，通过先验知识，I样本的增加为样本，因此通过经验风险最小化得到更可靠的。在图2（b）中，使用先验知识来约束，消除灰色区域以供以后优化。在图2（c）中，中最佳假设的θ的搜索策略被先验知识改变。特别是，虚线表示可以跳过经验风险的优化，这是元学习方法使用的策略。

因此，现有的作品可以被分类为一个统一的分类法，如图3所示。我们将在下一节详细介绍每个类别。

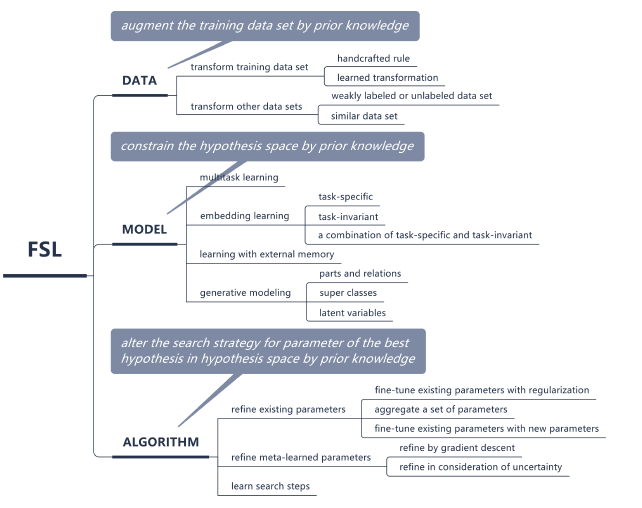


图3。基于每个方法的焦点的FSL分类。

**3.数据**

本节中的3种数据方法通过利用先验知识增加数据量来解决FSL问题，从而丰富了E中的监督信息样本复杂度需要后续的机器学习模型和算法，并获得更可靠的。

在这里，我们展示了如何使用先验知识在FSL中扩充数据。根据先验知识的类型，我们将这些方法分为四类，如表2所示。因此，图4展示了转换的工作原理。由于对中N个类中的每一个类的扩充是独立完成的，因此我们使用中N个类的示例进行说明。

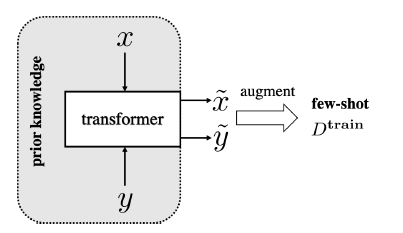


图4。通过装转换器的输出（,）变换某些输入（x，y））来增加数据集的说明。

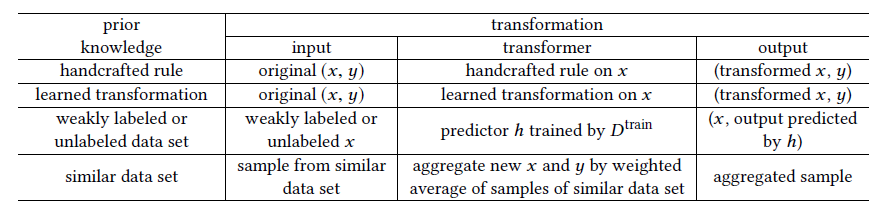


表2。聚焦于数据透视的FSL方法的特点。

**3.1转换**

此策略通过将每个转化为几个具有某些变化的样本来增强。转换过程可以从相似的数据中学习或由人类专家设计，作为先验知识包含在经验E中。它目前只应用于图像，因为合成的图像很容易被人评价。

**3.1.1手工规则**。在图像识别任务中，许多工作通过使用手工规则作为预处理例程来转换中的原始示例来增强，例如翻译[13、66、100、105]、翻转[89、105]、剪切[105]、缩放[66、143]、反射[31、63]、剪切[89、143]和旋转[100、127]给定示例。

**3.1.2习得转化**。与此相反，该策略通过将原始示例复制到多个示例中来增加，然后通过学习转换对这些示例进行修改。学习转换本身是E中的先验知识，而当前的FSL任务既不需要训练样本，也不需要学习过程。最早关于FSL的论文[78]通过迭代地将每个样本与其他样本对齐，从类似的类中学习一组几何变换。然后对每个应用该学习转换，形成一个可以正常学习的大数据集。同样，Schwartz等人。[2018]从一个类似的类学习一组自动编码器，每个代表一个类内变化，通过将变化添加到来生成新样本。假设所有类别在样本之间共享一般的可转换可变性，在[48]中学习了一个转换函数，通过类比将从其他类学习到的样本对之间的变化转移到。Kwitt等人没有列举成对内的变异性。[2016]使用从大量场景图像中学习的一组独立属性强度回归器，将每个转换为多个新样本，并将这些新样本分配给原始的标签。基于[65]，Liu等人。[2018]进一步提出学习一个连续的属性子空间，该子空间可用于将任何属性变化引入x。

**3.1.3讨论**。通过手工规则转换在deep模型中被广泛使用，以减少过度拟合的风险[41]。然而，深度模型通常是从大规模数据集中学习的，其中样本足以估计其粗糙分布（判别模型的条件分布或生成模型的分布）[80]。在这种情况下，通过增加更多的样本来增加可以帮助绘制更清晰的分布形状。相比之下，FSL只包含少量的监督信息，因此它的p（x，y）没有很好的暴露。这些手工编制的规则（如简单的缩放和旋转）可以转换所有图像，而不考虑中可用的任务或所需的数据属性。他们不会带来额外的监督信息。

因此，它们仅用作图像数据的预处理步骤。对于通过学习转换来转换，它是数据驱动的，利用了类似于任务T的的先验知识，因此可以增加更合适的样本。然而，这种先验知识需要从类似的任务中提取出来，这些任务可能并不总是可用的，而且收集起来成本很高。

**3.2转换其他数据集。**

此策略将来自其他数据集的样本转换为目标的，并使其与目标的相似，从而增强到监督信息数据集。

**3.2.1弱标记或未标记的数据集**。该策略使用大规模弱标记或无标记数据集。已知此数据集包含与相同标签的样本，但未显式给出输出。因此，我们必须首先找到这些带有目标标签的样本。

由于这些大规模数据集的样本包含样本的巨大变化，将它们增加到有助于描述更清晰的p（x，y）。考虑视频手势识别，Pfister等人。[2014]使用一个大而弱标记的手势库，其中包含不同人连续手势的大变化，但手势之间没有明显的间隔。从学习到的分类器用于从手势库中提取与手势相同的样本。然后利用这些样本建立最终的手势分类器。标签传播用于直接在[29]中标记未标记的数据集。

**3.2.2类似数据集**。此策略通过聚合来自其他相似但较大数据集的样本对来增强。例如，老虎的数据集类似于猫的另一个数据集。假设基础最优假设适用于所有类，类的x之间的相似性可以转移到类的y。因此，可以通过聚合来自相似数据集的样本来生成新样本，其中聚合权重通常是从其他信息源（例如文本语料库）提取的一些相似度量。但是，直接将聚合的样本增加到可能不合适，因为这些样本不是来自目标FSL类。因此，Gao等人。[2018]设计了一种基于生成性对抗网络（GAN）[42]的方法，从多个样本的数据集中生成任意合成的聚合。

**3.2.3讨论**。弱标记或未标记数据集的收集通常是廉价的，因为不需要人为地进行标记。然而，伴随着这种廉价性，这类数据集的质量通常较低，如数据集的粗糙和缺乏严格的收集和审查程序，导致合成质量不明确。此外，从这个大数据集中挑选有用的样本也很昂贵。类似的数据集与有一些共同的属性，并且包含了足够的监督信息，使其成为一个更具信息性的数据源。然而，为了寻找相似的数据集而确定关键属性是客观的，而收集这种相似的数据集是很困难的。

**3.3总结**

通过增加，本节中的方法能够满足期望的样本复杂度，并获得可靠的经验风险最小化。第一类变换方法通过手工或学习变换规则对每个原始样本进行∈变换。它们在原始样本的基础上增加了，因此构造的新样本不会离太远。但由于这个原因，以这种方式生成样本受到限制。第二类方法是对其他数据集的样本进行变换，使其适应于模拟∈。这些数据集是大规模的，为转换提供了大量变化的样本。然而，如何使这些样本变得像∈是很困难的。

一般来说，从增强的角度求解FSL是很简单的。考虑到合并问题的目标便于学习，可以对数据进行扩充。这种增强过程对人类来说通常是合理的。然而，由于p（x，y）未知，完全先验知识是不可能的。这意味着扩增过程并不精确。估计值与实际值之间的差距很大程度上影响了数据质量，甚至导致概念漂移。

**4模型**

模型确定由θ参数化的假设h（·；θ）的假设空间，以逼近从输入x到输出y的最优假设。

如果使用普通的机器学习模型来处理少量的数据训练，它们必须选择一个小的假设空间。如（5）所示，一个小的具有小的样本复杂度，因此需要训练的样本较少[80]。当学习问题简单时，例如特征维数较低时，一个小就可以获得期望的良好学习性能。然而，现实世界中的学习问题通常是非常复杂的，由于显著的，它们不能很好地用小中的假设h来表示。因此，大是FSL的首选，这使得一般的机器学习模型不可行。正如我们将在后续文章中看到的，本节中的方法通过E中的先验知识来补充样本的缺乏来学习一个大。具体来说，先验知识被用来影响的设计选择，例如约束H。这样，样本复杂度降低，经验风险最小化更可靠，并且减少了过度装配的风险。根据使用的先验知识，这类方法可进一步分为四类，如表3所

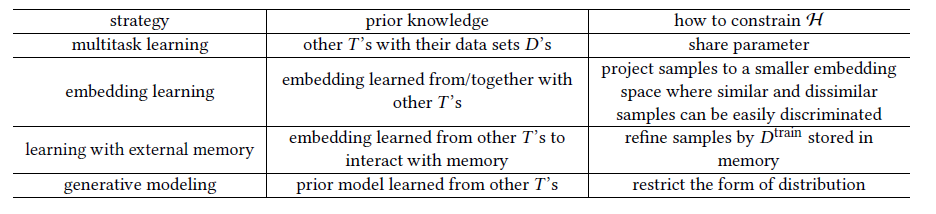


表3。聚焦于模型视角的FSL方法的特点。

**4.1多任务学习**

多任务学习[23]方法利用任务间共享的一般信息和每个任务的具体信息，自发地学习多个学习任务。它广泛应用于有限训练实例的多个相关任务并存的应用中，因此它们可以自然地应用于FSL问题。注意，当多任务学习处理来自不同领域的任务时，它也被称为领域适应[11]。

形式上，给定一组与R相关的任务，其中既包括少量样本的任务，也包括许多样本的任务，每个任务对数据集进行操作，其中由训练集和测试集组成。在这些任务中，我们称少数的shot任务为目标任务，而其余的任务为源任务。多任务学习从学习得到每个Tt的。由于这些任务是相关的，所以假设它们具有相似或重叠的假设空间。显然，这是通过在这些任务之间共享参数来实现的。这些共享参数可以看作是通过其他联合学习任务来约束每个的一种方法。根据参数共享是否被显式地执行，我们将该策略中的方法分为硬参数共享和软参数共享。关于硬参数和软参数共享的图示如图5所示。

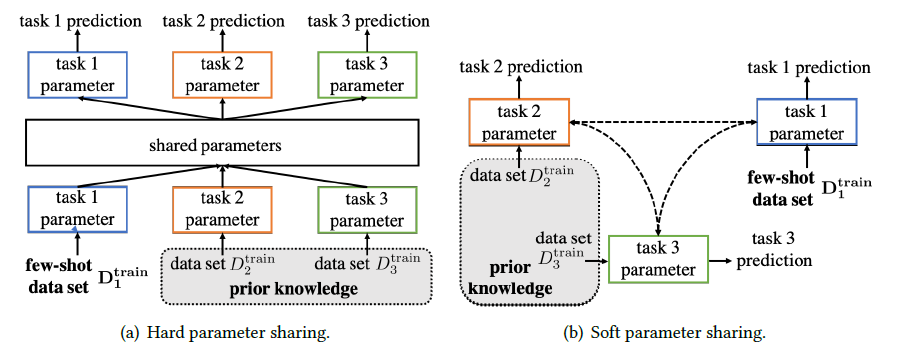


图5。软、硬参数共享策略在FSL问题多任务学习中的应用举例。

**4.1.1硬参数共享**。该策略明确地在任务之间共享参数，以促进重叠的，并且可以为每个任务学习特定于任务的参数，以考虑任务的特殊性。在[143]中，这是通过共享两个网络的前几层来学习一般信息，同时学习不同的最后一层来处理每个任务的不同输出来完成的。

Benaim和Wolf[2018]以相反的方式进行领域适应。它们学习不同域中源任务和目标任务的独立嵌入，将其映射到任务不变空间，然后学习共享分类器，对所有任务的样本进行分类。最后，文[81]中的方法首先从源域中的源任务预训练一个变分自动编码器，克隆它作为目标任务。然后它共享一些层来捕获一般信息，并让两个任务都有一些特定于任务的层。目标任务只能更新其特定于任务的层，而源任务可以更新共享层及其特定层。它避免了直接使用更新共享层，从而降低了过度拟合的风险。

**4.1.2软参数共享。**

此策略不在任务之间显式共享参数。相反，每个任务Tt都有自己的假设空间和参数。它只鼓励不同任务的参数相似，从而产生相似的结果，这可以通过正则化来实现。[2015]惩罚所有组合中的成对差异，迫使所有都被类似地学习。除了直接正则化外，另一种方法是通过损失调整来强制软参数共享。优化后，学习的也利用了彼此的信息。罗等人。[2017]通过预先训练的CNN从源域中的源任务学习，为目标域中的目标任务初始化CNN。在训练过程中，他们使用由CNN的多层表示计算出的对抗性损失来迫使两个CNNs将样本投影到一个任务不变空间。

**4.1.3讨论**。多任务学习方法通过一组共同学习的任务来约束每个任务的学习。通过显式或隐式地共享参数，联合学习的任务一起隐式地消除那些不可行的区域。通过硬参数共享很容易实现。共享的假设空间用于捕获共性，而每个任务都在其上构建特定的模型假设空间。相比之下，软参数共享只鼓励类似的假设，这是一种更灵活的约束的方法。但是如何实施相似性约束需要仔细设计。

**4.2嵌入学习**

嵌入学习[55，109]方法将∈⊆嵌入到一个较小的嵌入空间∈⊆中，在该空间中可以容易地识别相似和不同的对。因此，是受约束的。嵌入函数主要是通过先验知识学习的，还可以利用引入特定任务的信息。注意，嵌入学习方法主要是为分类任务设计的。

嵌入学习方法有以下几个关键组成部分：函数f（·）嵌入样本到，函数g（·）嵌入样本到，相似度量s（·，·）用于计算中每个的f（）和g（）之间的相似性。然后将分配给最相似的类。虽然g可以与f相同，但有时对和使用不同的f和g。这是因为可以根据来自的信息显式地嵌入，从而调整比较兴趣[15，127]。因此我们区分了这两个嵌入函数。通常，使用和的一组数据集Dc来学习这些组件。请注意，Dc可以是多个样本或少个样本的数据集。嵌入学习策略的示例如图6所示。我们还详细介绍了现有的嵌入学习方法，如表4中的f、g和s。

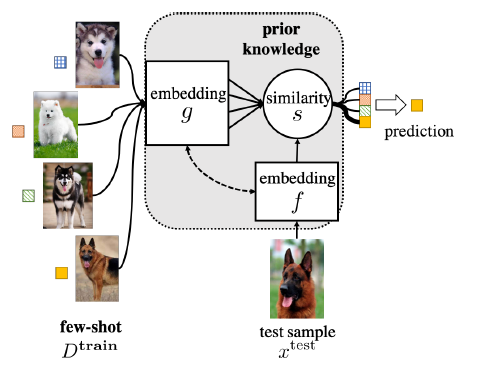


图6。FSL问题的嵌入学习策略说明。该图改编自[127]。

接下来，根据嵌入的信息，我们将这些方法分为任务不变量（换句话说，通用）、任务特定和两者的组合。

**4.2.1特定任务**。特定于任务的嵌入方法学习为D定制的嵌入函数。Triantafillou等人。[2017]学习嵌入以维护中每个的排名列表，其中同一类的排名更高或更低。由于使用了少量的镜头，因此通过将中示例之间的所有成对比较枚举为示例对，大大降低了示例复杂性。这样，每个原始示例都可以包含在多个样本对中，这在很大程度上降低了所需的样本复杂性。

**4.2.2任务不变量**。任务不变嵌入方法从一组不包含D的大数据集Dc中学习嵌入函数，假设这些嵌入能够成功地分离上的多个数据集，那么它们就足够通用，不需要再训练就可以很好地用于D。Fink[2005]提出了FSL的第一种嵌入方法。它从辅助Dc的核空间中学习，将和都嵌入到中，其中被赋给中的最近邻类。最近的一个深模型卷积连体网[61]学习孪生cnn将大数据集Dc中的样本对嵌入到公共嵌入空间Z中，然后使用的原始样本构造样本对，并将分类任务重新设置为验证/匹配任务，验证样本对的结果嵌入是否属于同一类。

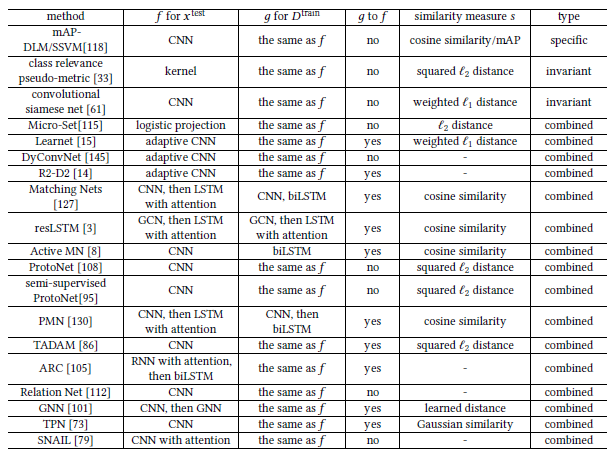


表4。嵌入学习方法的特点。“g to f”一栏表示f直接受如何嵌入“g”的影响。

**4.2.3任务不变量和特定任务的组合**。任务特定的嵌入方法充分考虑了任务的特殊性，而任务不变的嵌入方法可以在不需要重新训练的情况下快速地推广到新的任务。一种趋势是结合上述方法的优点：学习利用中包含的特定任务信息来适应从先验知识中学习到的通用任务不变量嵌入空间。Tang等人。[2010]首次提出以微集的名义优化FSL任务的过度分配。他们通过逻辑投影从这些FSL任务中学习Z。

对于一个新的少镜头任务T，将和中的所有样本映射到Z上，然后利用Z上的最近邻分类器对进行分类。

最近的研究主要使用元学习方法来融合任务不变量和特定任务的知识。我们根据核心思想对其进行分组，并突出代表性作品

（1） Learnet[15]在卷积暹罗网[61]的基础上，结合了每个任务T到Z的的特点，改进了卷积暹罗网[61] 。它从的元训练数据集学习元学习者将映射到卷积连体网络中各层的参数。为了减少学习者的参数数量，DyConvNet[145]使用了一组固定的过滤器，并且只学习为学习者组合它们。最近的工作[14]用岭回归模型代替Learnet的分类层，岭回归模型的参数可以用廉价的闭式解找到。

（2） 匹配网[127]将赋给Z中最相似的，其中和被f和g不同地嵌入。特别地，f是以为条件的，并且用双向LSTM（biLSTM）[45]聚集中所有示例的信息。然而，双向LSTM隐式地强制执行中示例之间的顺序。由于存在消失梯度问题，相邻算例之间的相互影响较大。为了消除不自然的顺序，Altae Tran等人。[2017]用LSTM代替了在其中使用的biLSTM，并进一步迭代地细化了其中的g和f来编码上下文信息。[8]中的主动学习变体将样本选择阶段添加到匹配网络[127]中，匹配网络[127]可以标记最有利的未标记样本，并使用它来增强。

（3） ProtoNet[108]只对中的和每个类的prototype进行一次比较。这个n类的原型被定义为该类嵌入的平均值，即其中是中n类的K个例子之一。ProtoNet使用同一CNN嵌入和，忽略了不同的特殊性，注意到[130]中提出了最佳匹配网和ProtoNet的结合，以考虑特定任务的信息。此外，Oreshkin等人。[2018]以平均cn为任务嵌入，然后映射到ProtoNet中使用的CNN的一些参数。文[95]提出了ProtoNet的一个半监督变量，它学习在学习过程中软分配相关的未标记样本以增强。

（4） 相关表示进一步将和由计算得到的每个cn的嵌入联合嵌入到Z中，然后将其直接映射到相似度评分分类中。这一思想在ARC[105]和relation net[112]中得到了独立的发展。ARC使用一个RNN，注意重复比较和每个类原型cn的不同区域并产生相对表示，另外使用一个biLSTM嵌入其他比较的信息作为最终嵌入。关系网首先使用一个CNN嵌入和到Z，然后将它们连接起来作为相对表示，并由另一个CNN输出相似度得分。

（5） 关系图是保持样本之间所有成对关系的图。具体地说，该图是以和的样本作为节点构造的，节点间的边由学习到的s确定，然后利用邻域信息对每个进行预测。文[101]中用GCN来学习中的与中的之间的关系图。节点的结果嵌入用于预测。相比之下，Liu等人。【2019】元学习嵌入，将和映射到Z，在Z上建立关系图，并根据闭式标签传播规则对进行标签。

（6） SNAIL[79]设计了由交织时间卷积层和注意层组成的特殊嵌入网络。时间卷积用于聚合来自过去时间步的信息，并且注意力选择性地关注与当前输入相关的特定时间步。网络参数是跨任务的元学习。在每一个任务中，网络依次取∈，并立即预测。

**4.2.4讨论**。任务特定嵌入充分考虑了D的领域知识，然而，给定的少量镜头数据可能会有偏差，只有从中学习可能是不合适的。在中建模排名表存在着与过度拟合的高风险。此外，这种学习方法不能从新任务中概括出来，也不容易适应。使用预先训练的任务不变量嵌入有计算成本低。然而，学习的嵌入函数不考虑任何特定于任务的知识。特别是由于只有少数几个例子，如针对少数情况的学习，单纯应用任务不变量嵌入函数是不合适的。任务不变量和任务特定信息的组合通常是通过元学习方法来学习的。它们能为学习者提供一个良好的，并能快速地概括出不同的任务。然而，如何在不引入负迁移的情况下，对一个新的、无关的任务进行概括，目前还不确定。

**4.3外部记忆学习**

外部存储器，如神经图灵机（NTM）[46]和存储器网络[111，133]允许短期记忆和基于规则的操作[46]。注意，学习是将训练样本的有用信息映射到模型参数的过程。给定训练数据集为的新任务T，必须重新训练该模型以合并其信息，这是代价高昂的。相反，使用外部记忆进行学习直接将所需的知识存储在外部记忆中以供检索或更新，因此它减轻了学习的负担，并允许快速概括。形式上，将存储器表示为，其中有b个存储器槽。

给定一个样本，它首先被f嵌入为查询q=f（）∈，然后关注每个，i=1。，b通过一些相似性度量s，例如余弦相似性。然后利用相似度确定从记忆中提取的知识，并据此进行预测。表5介绍了每个具有外部存储器的方法的详细特性。

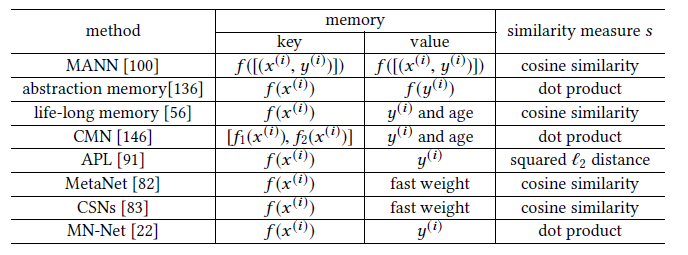


表5。外部记忆策略学习的特点。f是预先训练的嵌入函数，通常是CNN或LSTM。f1和f2是不同的嵌入函数。

对于FSL，样本有限，重新训练模型是不可行的。使用外部存储器进行学习可以将从中提取的知识存储到外部存储器中，从而帮助解决这个问题。从先验知识中学习的嵌入函数f没有重新训练，因此初始假设空间没有改变。当一个新的样本出现时，相关的内容会从内存中提取出来，并合并到这个样本的局部近似值中。然后将近似值反馈给后续模型进行预测，并对模型进行预训练。由于存储在内存中，因此可以有效地使用特定于任务的信息。总之，这种方法通过存储在内存中的来细化和重新解释样本，从而重塑。图7显示了嵌入学习策略的示例。

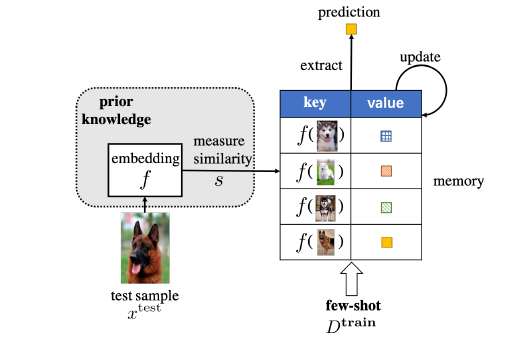


图7。FSL问题的外部记忆学习策略说明。

通常，当内存未满时，可以将新样本写入空内存插槽。

但是，当内存已满时，必须决定哪些内存插槽要更新或由某些设计的规则替换。我们根据这些更新规则中显示的不同偏好对现有作品进行分组，如下所示。

（1） 更新最近最少使用的内存插槽。最早的工作[100]使用内存来解决FSL分类问题，MANN[100]在内存已满时为新样本更新最近使用最少的内存插槽。当图像标签绑定在多个任务之间进行无序处理时，MANN更关心将同一类的样本映射到同一个标签。反过来，同一类的样本一起优化它们保存在内存中的类表示。

（2） 通过基于位置的地址更新。一些工作使用NTM中提出的基于位置的寻址，它通过梯度的反向传播随时更新所有的内存插槽。抽象存储器[136]使用这种更新策略。在每个任务中，元学习者首先从包含大规模辅助数据的存储器中提取相关的f（[]），然后将它们发送到抽象存储器。抽象内存的输出用于预测。

（3） 根据内存插槽的使用年限进行更新。有些内存记录每个内存插槽的使用年限。内存插槽在读取时会增加其使用期限，在更新时会将使用期限重置为0。最古老的更像是过时的信息。终身存储器[56]和CMN[146]在存储器满时都会更新最旧的内存插槽。然而，有时一个值在旧内存插槽中的罕见事件。为了解决这个问题，终身内存特别喜欢更新同一类的内存槽。由于每个类都占用相当数量的内存槽，因此很少有类受到某种方式的保护。

（4） 只有当丢失率很高时才更新内存。基于惊讶的存储器模块[91]设计了一个存储器更新规则，该规则仅在某些的预测损失超过阈值时更新存储器。因此，与可微存储器相比，该存储器的计算成本降低，并且存储器包含用于等价预测的最小但不同的信息。

（5） 将内存用作存储而不进行更新。MetaNet[82]将∈的样本级快速权值存储在内存中，并通过提取的快速权值对其进行嵌入和分类，从而将泛型信息和特定信息结合起来。MetaNet反复将快速权重应用于CNN的选定层。相比之下，Munkhdalai等人。[2018]学习快速权重，改变每个神经元的激活值，计算成本更低。

（6） 将新信息聚合到最相似的信息中。MN Net[22]将新样本的信息合并到其最相似的内存槽中。在匹配网络[127]中，内存不是直接预测，而是用于优化f（），并将CNN参数化为Learnet[14]。然后每个被这个条件CNN嵌入，并通过最近邻搜索与f（）匹配。

**4.3.1讨论**。只需将放在内存中就可以适应新的任务，在内存中可以很容易地进行快速的泛化。此外，终身学习或减少记忆更新等偏好也可以纳入到记忆更新和访问规则的设计中。然而，它依赖于人类的知识来设计所需的规则。现有的作品没有明确的赢家。如何根据不同的设置自动设计或选择更新规则是一个重要的问题。

**4.4生成建模**

这里的生成建模方法是指涉及p（x，y）的方法。他们利用先验知识和来获得估计的分布。先验知识通常是从由训练集和测试集组成的一组C数据集Dc中学习到的一些概率分布参数表示的先验模型。通常情况下，数据中心很大，数据中心不属于数据中心的一部分。生成性建模方法更新数据中心的概率分布以进行预测。图8展示了这种策略。

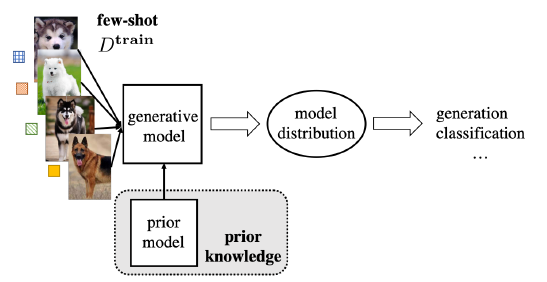


图8。FSL问题的生成性建模策略说明。

具体地说，对于给定的，的后验概率y由Bayes规则计算为，即。通过参数化展开，可以写成，其中是h∈的参数。如果足够大，我们可以用它来学习一个很好的峰值，并使用最大似然估计（MLE），或最大后验（MAP）。

然而，FSL任务中的样本有限，不足以学习。因此，它不能学习一个好的θ。

FSL的生成模型假设θ在不同的y（例如，类）之间是可转移的。因此，θ可以从一组大的数据集Dc中学习。具体地说，将扩展为，它也有参数γ。然后利用自适应分布或学习γ得到。

表6总结了该策略中的主要参考文献及其特点。通过从先验知识中学习先验概率，限制了的形状。根据θ在y上的定义和共享，我们将现有的方法分为部分和关系、超类和潜在变量。

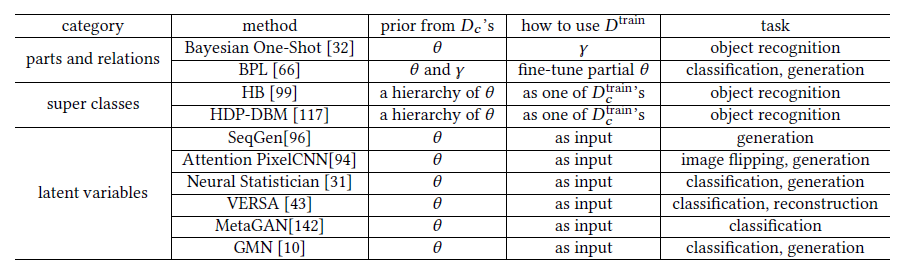


表6。生成性建模方法的特点。

**4.4.1零件和关系**。该策略从大量Dc的先验知识中学习部分和关系（即a.k.a.）。尽管目标shot类很少有示例，但它们的组件（如parts和relations）与许多类共享。使用更多的示例，部件和关系更容易学习。对于∈，模型需要推断相关部分和关系的正确组合，然后确定该组合属于哪个目标类。Bayesian One Shot[32]和BPL[66]属于这一类。Bayesian One Shot利用对象的形状和外观来帮助识别对象，而BPL将字符分为类型、标记和进一步的模板、部件、原语来建模字符。Bayesian单发中使用了少量的部分，大大减少了部分和关系的组合空间，而BPL只考虑了五种最可能的组合。

**4.4.2超类**。部分和关系转化为小部分样本的模型，而超级类则通过无监督学习对相似类进行分组。在考虑分类任务的情况下，该策略将这些超类的参数化h∈的最优θ作为先验知识。一个新的类首先被分配给一个超类，然后通过调整超类的h来学习它的h。在[99]中，他们学习使用Dc（包括D）来形成类的层次结构。这样，相似的类一起有助于学习一个精确的表示超类的一般先验，而作为回报，每个超类都可以为其指定的类提供指导，特别是为数不多的示例的提供指导。在[117]中，通过使用深Boltzmann机器来学习更复杂的特征，进一步改进了[99]的特征学习部分。

**4.4.3潜在变量**。将样品分成若干部分和相互关系是手工制作的，并且在很大程度上依赖于人类的专业知识。相反，这种策略对潜在变量进行建模，没有在类之间共享的隐式含义。在不进行分解的情况下，从Dc学习的h不再需要调整，从而大大降低了新任务的计算成本。为了处理更复杂的，该策略中使用的模型通常是深度模型。根据经典的深层生成模型作为基础，现有的作品可以分为以下几类。

（1） 变分自动编码器（VAE）[59]。Rezende等人。[2016]建议使用序列VAE通过一组序列推断的潜在变量对p（x）建模。该模型反复关注中每个的不同区域，分析了当前模型提供反馈的能力，能够很好地对密度估计进行建模。

（2） 自回归模型[123]。里德等人。[2018]提出了一个自回归模型，将图像的密度估计分解为像素级。该模型根据已经生成的像素和从存储的存储器中获取的相关信息顺序地生成每个像素。

（3） 推理网络[141]。Edwards和Storkey[2017]学习了一个推理网络来推断这些潜在变量，以及另一个推理网络来将映射到θ，θ是其生成分布的参数。同时，Gordon等人也通过分期变分推理学习推理网络。[2019]学习将映射到变分分布的参数，该变分分布近似于输出y上的预测后验分布。

（4） 生成性对抗网络（GAN）[42]。Zhang等人。[2018b]利用分类任务的判别模型联合学习不完全GAN。不完全GAN产生的样品与中的样品相似，但略有不同。通过学习区分和稍有不同的伪数据，得到了更清晰的决策边界。

（5） 匹配网的生成版本（GMN）[127]。Bartunov和Vetrov[2018]将区分匹配网络扩展到生成性设置，即GMN。GMN将匹配网络中使用的替换为一个潜在变量z，然后将嵌入其中，将z嵌入到嵌入空间z中，其中f（z）参与到每一个ρ（）中，以获得聚集的权重。

**4.4.4讨论**。通过将每个对象分解成更小的部分来学习它们，关系利用人类的知识来进行分解。与讨论的其他类型的生成建模方法相比，零件和关系更易于解释。然而，人类的知识具有很高的偏差，这可能不适合给定的数据集。此外，它可能很难获得，也可能很昂贵，对应用程序场景设置了严格的限制。相比之下，为超级类学习的模型可以聚合来自许多相关类的信息，并充当新类的通用模型。然而，它们可能不是最优的，因为没有使用特定的信息。具有潜在变量的方法效率更高，对人类知识的依赖更少。然而，由于潜在变量的确切含义未知，这类方法不容易理解。

**4.5总结**

本节中的所有方法都是基于经验E中的先验知识来设计，以限制的复杂性并降低其样本复杂性。

多任务学习方法通过一组联合学习任务来约束少量任务的H。

它们可以在不同的任务之间进行通信，并在优化过程中改进这些任务。它们还隐式地增加数据，因为一些参数是由多个任务共同学习的。然而，目标D必须是之一才能进行联合训练。因此，对于每一项新任务，都必须从头开始学习，这可能是一项代价高昂、速度缓慢的任务。它不适用于只有一次机会或偏好快速推理的任务。

嵌入学习方法学习将样本嵌入一个较小的嵌入空间，在这个空间中可以很容易地识别相似和不同的对。因此，是受约束的。大多数作品从大规模数据集中学习任务不变信息，并能吸收新任务的任务特性。一旦学习，大多数方法可以通过前向传递从新任务中简单地进行归纳，并在嵌入的样本中执行最近邻。然而，如何将θ内任务的不变信息和特定信息进行原则性的混合还不清楚。

外部记忆学习方法通过存储在内存中的对每个样本进行细化和重新解释，从而重塑。通过显式地将存储在内存中，避免了为适应而进行的费力的重新训练。任务专用信息得到有效利用，不易遗忘。然而，使用外部存储器进行学习会产生额外的空间和计算成本，这会随着存储器的增大而增加。因此，当前的外部存储器容量有限，因此不能存储很多信息。

生成性建模方法从先验知识中学习先验概率，形成的形式，具有很好的解释能力、因果性和综合性[66]。通过学习联合分布p（x，y），他们可以处理更广泛的任务，如生成和重建。学习的生成模型可以生成多个样本进行数据扩充。然而，生成性建模方法通常具有较高的计算成本，与其他模型相比很难推导。为了计算的可行性，它们需要对结构进行严格的简化，从而导致不精确的近似。

**5算法**

算法是在假设空间中搜索最佳假设的参数θ的策略。例如，随机梯度下降（SGD）及其变体[19，20]是一种流行的搜索的策略。在SGD中，θ通过一系列更新t=1，….来更新。在第t次迭代中，设，然后由



其中是要调谐的步长。当监督信息丰富时，有足够的训练样本来更新到任意精度，并通过交叉验证找到合适的。然而，所提供的少量的不足以达到所需的样本复杂度。因此，所得到的经验风险最小化方法是不可靠的。

本节中的方法不限制的形状，因此仍然可以使用常见的模型，如CNN和RNN。相反，他们利用先验知识来改变对的搜索，参数化中的来解决FSL问题。根据搜索策略如何受到先验知识的影响，我们将本节中的方法分为三类（表7）：

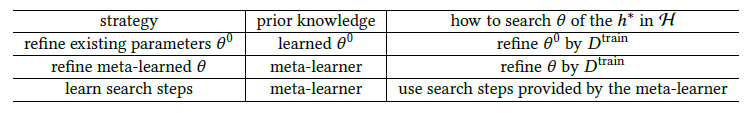


表7。基于算法视角的FSL方法的特点

（1） 优化现有参数。从其他任务中学习的初始用于初始化搜索，然后由进行优化。

（2） 优化元学习。元学习者从一组与少镜头任务相同的任务分布中学习出一个一般θ，然后每个学习者使用细化元学习者提供的θ。

（3） 学习搜索步骤。此策略学习元学习者输出搜索步骤或更新规则以直接指导每个学习者。它不是学习更好的初始化，而是改变搜索步骤，例如方向或步长。

**5.1完善现有参数**

该策略以预训练模型的作为良好的初始化，并通过对其进行自适应。假设通过从大规模数据中学习来捕获一般结构，因此可以使用少量迭代来调整它，以便在D上很好地工作。

**5.1.1正则化微调**。这种策略通过一些正则化来微调给定的。图9展示了这种策略。微调在实践中被广泛使用，它通过反向传播将在大规模数据（如ImageNet）上训练的（deep）模型的值调整为较小的数据集[27]。包含泛型知识的单θ0通常是深部模型参数化了一个庞大而复杂的，在有限的条件下，通过梯度下降对进行简单的微调会导致拟合过度。如何在不过度拟合的情况下适应的取值是设计的关键问题。

在本节中，方法使用正则化微调以防止过度拟合。它们可以按如下方式分组。

（1） [6]中使用了提前停车。然而，它需要一个独立于的验证集来监控训练，这进一步减少了训练样本的数量。此外，使用一个小的验证集使得搜索策略有很大的偏差。

（2） 选择性地更新是指仅更新的一小部分以避免过度拟合。Keshari等人。[2018]使用一组固定过滤器，只学习通过安装Dtrain来控制过滤器中的大量元素。如果接受了预先训练的CNN，乔等人。[2018]和Qi等人。[2018]直接在最后一层的权重矩阵中添加中每个n类的权重作为新列，同时保留预先训练的权重不变。

（3） 聚类和使用相同的更新信息更新得到的参数组可以在很大程度上约束搜索策略。在[139]中，他们使用辅助数据对预先训练的CNN的文件管理器进行分组，并使用通过分组反向传播对CNN进行微调。

（4） 模型回归网络[132]假设存在一个任务无关的转换，从使用几个示例训练的参数到使用许多示例训练的参数。Wang和Hebert[2016b]然后用固定的N-way-K-shot问题改进学习。类似地，Kozerawski和Turk[2018]学习将的嵌入转换为分类决策边界。

**5.1.2聚集一组**。通常，我们没有合适的θ进行微调。相反，我们可以从相关任务中学习许多模型参数，例如任务是人脸识别，而只有眼睛、鼻子、耳朵的识别模型可用。因此，可以从θ0中选取相关的值，并将它们聚合到合适的初始化中，以供适应。图10展示了这种策略。

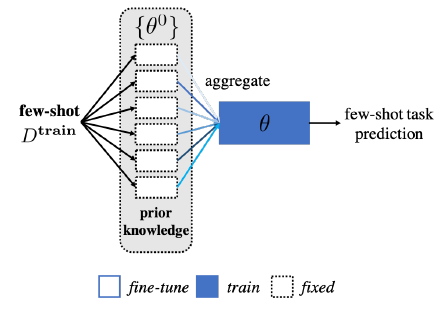


图10。将一组聚合成θ的策略的说明。

这里的参数通常是从其他数据集中预先训练的。根据所使用的数据集，现有方法可以按如下方式分组。

（1） 类似的数据集。Bart和Ullman[2005]根据图像片段对一类图像进行分类。新类的分类器是通过用新类的相似特征替换已学习类的特征并重用它们的分类器参数来构建的。只调整分类的阈值，以避免与那些类似的类混淆。类似于[89]和[90]，一个预先训练的CNN被改编成处理[40]中的新课程。但是它并没有仅仅使用的嵌入作为高度偏倚的分类器参数，而是构造这个新类的分类器是通过嵌入的线性组合和通过考虑其他类的分类器参数而建立的试探性分类器。

（2） 未标记的数据集。从未标记的数据集中学习的对于分离样本也是有区别的。对未标记数据集中的样本迭代地分配和调整伪标签，以便学习决策边界[131]。此外，这些学习到的决策边界被合并到预先训练的CNN中[131]。注意，在预先训练的CNN中，嵌入层越高，它们就越具体。通过学习分离未标记的数据集，提高了最后一层嵌入的通用性。具体来说，给定一个预先训练好的CNN，Wang和Hebert[2016a]在完全连接层前添加一个特殊的层进行分类，固定其余的预先训练参数，然后学习新的γ，可以很好地分离未标记的数据集。要对新任务进行分类，只需学习线性层进行最终分类，并重用CNN的其余部分。

**5.1.3用新参数微调**。预先训练的可能不适合新的FSL任务的结构。因此，我们需要为的特性增加新的参数δ。具体来说，该策略在学习δ的同时微调，使得要学习的模型参数变成θ={，δ}。图11展示了这种策略。霍夫曼等人。[2013]使用预先训练的CNN的下层参数用于特征嵌入，而在其上使用学习线性分类器。考虑字体样式转换任务，Azadi等人。[2018]预先训练一个网络来捕获灰色图像中的字体，并与生成时尚彩色字体的网络训练一起进行微调。

**5.1.4讨论**。本节中讨论的方法减少了从头开始对H进行架构搜索的工作量。由于直接微调容易过拟合，用正则化微调一个的方法转向正则化或修改现有参数。他们通常考虑某个深模型的单个。然而，合适的现有参数并不总是容易找到的。另一种方法是将相关任务中的一组参数聚合到适当的初始化中。但是，必须确保这些现有参数中嵌入的知识对当前任务有用。此外，搜索大量现有参数以找到相关参数的成本也很高。使用新参数微调可获得更大的灵活性。但是，考虑到很少的快照，只能添加有限的参数，否则可能会出现过度拟合。

**5.2细化元学习θ**

在以下章节中均为元学习方法。这一战略不是朝着不可靠的方向努力，而是直接针对。在下面，我们表示元学习者的参数为θ，元训练任务的任务特异性参数Ts为，元测试任务Tt为。在训练过程中，由θ参数化的元学习者（优化器）为任务Ts（优化器）的学习者提供信息，学习者将梯度等错误信号返回给元学习者以改进它。然后，给出一个带有的元测试任务Tt，可以直接使用元学习者，而学习者可以从中学习。我们主要说明使用元测试任务Tt。图12展示了这种策略。

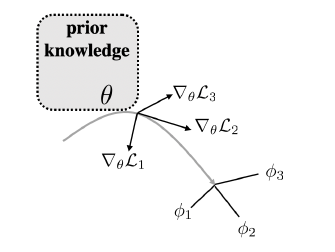


图12。改进元学习θ策略的实例。该图改编自[34]。

**5.2.1通过梯度下降进行细化**。该策略通过梯度下降来改进元学习θ。模型不可知元学习（MAML）[34]是这类方法的代表。对于任务，它获得一个θ作为良好的初始化。可以通过使用的一些梯度下降来有效地调整该，以获得良好的特定于任务的ˍTs。从数学上讲，这是通过来完成的，其中

和α是要选择的固定步长。通过对所有样本的求和，它提供了一个置换不变量。

然后，元学习者通过所有元训练任务的平均梯度步长更新θ，，其中，β也是固定步长。

MAML为所有任务提供相同的初始化，而忽略特定于任务的信息。这只适用于一组非常相似的任务，而当任务是不同的时，效果很差。在[69]中，它学习从θ的一个子集中为新Tt选择初始值。换句话说，它元学习任务特定的子空间和度量，以便学习者执行梯度下降。因此，对不同的Tt提供了不同的θ初始化。

由于梯度下降法可能不可靠，因此采用正则化方法来校正下降方向。由模型回归网络[132]在[47]中进一步调整。调整后的被正则化为更接近用许多样本训练的模型。模型回归网络的参数是通过类似θ的梯度下降来学习的。

**5.2.2考虑到不确定性进行改进。**用几个例子学习必然会产生一个具有高度不确定性的模型[36]。所学习的模型能以很高的置信度预测新任务吗？这个模型会用更多的样本来改进吗？测量这种不确定性的能力为主动学习或进一步收集数据提供了一个标志[36]。

因此，到目前为止考虑的不确定性有三种。

（1） 共享参数θ的不确定性。单个θ不能作为所有任务的良好初始化。因此，通过对θ的后验分布建模，可以为不同的Tt’s采样适当的初始值。[2018]提出对θ的先验分布进行建模。它是通过地图的点估计来解决的。相反，Yoon等人。[2018]通过Stein变分梯度下降（SVGD）学习θ的先验分布。他们学习了一组θ的拷贝用共享参数为。这些副本相互通信以确定更新方向。使用所学的，通过几个SVGD步骤获得。

（2） 任务特定参数的不确定性。然后，每个是上后部的点估计值。然而，由于包含一些示例，因此学习到的后验可以是倾斜的。因此，Grant等人。【2018】通过用拉普拉斯近似代替后验点估计来改进MAML。Ravi和Beatson[2019]不使用θ作为。相反，他们学习推理网络，将映射到。然后，如在MAML中一样，通过使用采取几个梯度下降步骤来优化该推理网络，以便使适应。最后，使用作为变分分布的参数，该变分分布近似于上的后验分布。

（3） n类特定参数的不确定度，。最后，在[98]中对特定类别的不确定度进行了建模。它首先将中的第n类中的每一个映射到低维潜在空间中的类条件多变量高斯的参数，以便对依赖于类的初始化进行采样，然后使用编码器嵌入类n到中的那些，其中，梯度下降被有效地取到了关于的值， 最后，使用解码器将学习到的类相关参数映射到另一类条件多元高斯的参数，该类条件多元高斯用于采样。

**5.2.3讨论**。在本节中，要优化的θ是从任务分布中提取的一组任务的元学习。如果新任务属于训练元学习者的任务分布，则使用元学习θ作为新任务的初始化更为合适。

然而，现实情况可能并非如此。在考虑不确定性的情况下，对θ进行细化，可以度量新任务预测的置信度，为主动学习或进一步的数据采集提供信号。然而，学习建模这种不确定性的成本要高得多，初学者很难进行设计。

**5.3学习搜索步骤**

上一节中使用的元学习者为每个h提供一个良好的初始化θ，而在这里学习元学习者直接为每个学习者输出搜索步骤或更新θ在h的规则[5，70]。它可以依次取与学习者互动。这样，无需调整步长α或寻找最佳下降方向，学习算法自动完成。图13展示了这种策略。

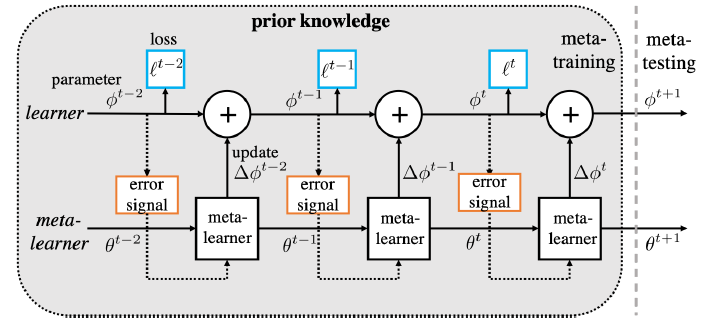


图13。学习搜索步骤策略说明。该图改编自[5]。

Ravi和Larochelle[2017]首先提出通过学习H中的搜索步骤来解决FSL问题，他们改进了Andrychowicz等人的工作。[2016]，我们将介绍为基础。在训练过程中，θ不会在每个任务Ts中更新。在第t次迭代中，元学习者使用第t次样本和在t-1次迭代中学习的计算出，并通过将输入θ参数化的坐标LSTM，生成更新的。注意，对应于和（6）中梯度的组合。最后，它将更新为。当学习结束时，对于一个N-way-K-shot分类任务，利用反向传播梯度下降法在上测量的损失来提高θ。

然后，通过从P（T）中提取一组T s来改进元学习者，它在提出有效的FSL算法方面变得越来越好。基于[5]，除了避免调整α之外，Ravi和Larochelle[2017]还为ɕTs提供了良好的初始化。它们通过在LSTM内通过单元更新来实例化（6）而直接获得。

**5.3.1讨论。**学习搜索步骤方法改变了中的θ的搜索策略。通过从一组任务中进行元学习，元学习者捕获了这类任务的常用搜索策略。当它通过先验知识进一步了解时，它可以建议更好的搜索方向或速度，如果仅基于每个任务的进行搜索，则可以稍后考虑。

**5.4总结**

本节中的方法通过设计适当的算法来解决FSL问题。具体来说，它们利用先验知识来搜索参数化的θ。先验知识可以通过提供一个好的开始搜索的初始点或直接提供搜索步骤来改变搜索。

改进现有的参数方法减少了架构搜索H或从头学习的工作量。利用已有的作为初始值，学习通常需要较少的计算量和迭代次数来获得一个好的假设h∈，这种策略保留了训练好的模型的参数而不是原始数据，使得知识的迁移具有可伸缩性。然后，学习目标是优化这些现有参数。由于是从他们自己的任务而不是当前的T中学习的，因此学习如何使这些参数适应T对于目标是间接的，并且很容易误入歧途。换句话说，这种策略通常为了方便而牺牲精确性。

另外两种策略都是元学习策略，元学习参数θ是学习的良好初始化，学习的搜索步骤直接用作的搜索步骤。通过对一系列相关任务的学习，元学习θ更接近于新任务Tt的最终任务特定参数。然后，学习搜索步骤通过使用元学习者来处理学习过程中的整个算法部分。但是，每个任务的初始化都受以前任务的影响，而不检查连续任务的相关性。一旦任务是负相关或完全不相关的，这种元学习方法就不能提供良好的初始化，并且容易产生负迁移。需要注意的是，考虑不确定性的改进方法只能释放出不合适的信号，而不能解决负迁移问题。

**6未来工作**

FSL方法将先验知识与E中提供的少量监督信息相结合，使监督目标的学习成为可能。本文从问题设置、技术、应用和理论四个方面论述了FSL进一步发展的关键方向

**6.1问题设置**

现有的FSL方法使用单一模态的先验知识，如图像、文本和视频。虽然在当前的模式中有一些例子，但是可能存在一个具有大量监督样本的更便宜的模式。例如，一个灭绝的动物类在视觉领域。然而，它可以在文本领域，如教科书或网页，详细审查，因为人们往往特别注意罕见的课。因此，可以利用多模态先验知识在互补视图中提供先验知识。零拍学习中经常使用多模态的思想。他们考虑了先验知识，包括属性[1，54]、WordNet[1，54]、单词嵌入[121，128]、共现统计[77]和知识图[129]。

近年来，在中，有一些工作尝试采用零炮学习方法来处理FSL问题。这些方法要么使用少量的来微调由零炮学习方法[1，54]学习的参数，要么强制嵌入或由多个模式学习的模型在共享空间上匹配[121，128]。这样的微调可能会导致过度拟合，并且可能不会在参数中嵌入足够的新任务信息。例如，由于形式可能有不同的结构，文本必须服从句法结构，而图像则不服从句法结构。强迫他们配对可能会使学习变差。一个很有前途的方向是将多模态作为一种先验知识，并在数据、模型和算法的基础上进行设计处理。

**6.2技术**

现有的FSL方法分为数据（第3节）、模型（第4节）和算法（第5节）。事实上，为了改进这些方法，这些方法的每一个组成部分都可以被更新和更先进的方法所取代。例如，使用ResNet[50]作为嵌入函数可以比使用VGG[110]更好。此外，设计一种混合方法，可以显式地量化数据、模型和算法的贡献，并对方法进行相应的调整。

特别注意元学习方法，我们在第4节和第5节都讨论过。通过跨任务学习，可以快速适应新任务，推理代价小。然而，元学习方法中考虑的任务主要来自单个任务分布p（T）。然而，在实际工作中，由于成本太高或难度太大而无法确定任务相关性的任务有很多。在这种情况下，直接使用这些任务会导致负迁移[87]。最后，他们主要考虑静态和固定P（T）[34，93]。然而，p（T）通常是动态的[35]，新任务不断地以流的顺序出现。p（T）应该包含这种变化，另外，如何避免灾难性遗忘[60]对于这种动态p（T）是另一个问题，这意味着关于过去任务的信息不应该被遗忘。

自动机器学习（AutoML）[138]在有限的计算预算内，无需人工辅助就可以构建机器学习程序。现有的FSL方法各有利弊（如前几节所述），并且没有绝对的赢家为所有设置。除此之外，假设空间和搜索策略在人类设计中仍然起着重要的作用。一个可能的方向是将AutoML的自动特征工程[57]、模型选择[62]和神经架构搜索[148]扩展到FSL方法。因此，通过使此过程自动化，可以使人类从针对不同设置的手动设计方法中解放出来。

**6.3应用**

现有的工作主要涉及计算机视觉应用，例如字符识别[15，33，34，56，61，82，99，100，105，108，118，127，134]和图像分类[34，61，82，93，105，108，115，118，120，127，131，132，136]。这是因为视觉信息容易获得，并且在机器学习中得到了广泛的研究。有许多成熟的技术需要转移到很少的镜头设置。此外，视觉信息的表现可以很容易地被人类理解和评价。目前，字符识别和图像分类的两个基准Ominiglot和minimagenet已经得到了很高的精度几乎没有改进的余地[119]。文献[119]最近提出了一个由不同数据源构建的大规模多样性数据集，作为一个新的基准。此外，还可以探索更多的计算机视觉应用，例如图像检索[118]、目标跟踪[15]、手势识别[88]、图像字幕和视觉问答[28]以及视频事件检测[137]。

除了计算机视觉的应用外，很少有镜头问题开始在其他领域被考虑。在自然语言处理中，例如翻译[56]和语言建模[127]。

在建议中，冷启动项目建议已在[126]中得到考虑。在医学应用中，文献[3]很少尝试发现注射药物。[21]中考虑了一次性架构搜索。最后，机器人和游戏的强化学习[113]从新环境中有限的经验开始引起人们的注意[30，34，79]。现有的应用是单次模拟[30]、多臂匪徒[30]、视觉导航[30]、运动中的连续控制[34]。

最近，这些应用进一步扩展到动态环境[284]。

**6.4理论**

FSL利用先验知识来弥补监督信息的不足。基本上，先验知识的使用降低了样本复杂度，这在使用无标记数据集[12，107]、多任务学习[85]、生成建模[39]、优化现有参数[76]和元学习[4]的方法中得到了证明。但是，对于条件嵌入学习和外记忆学习等新的学习方法，目前还缺乏分析。Finn和Levine[2018]表明，使用足够深入的模型，MAML[34]可以近似任何假设h。除此之外，回忆性FSL问题有时会成为领域适应问题，如第4.1节所述。现有的域自适应界可能是鼓舞人心的[11，18]，即，在[76]中示出了通过传递具有微调的表示而产生的前馈神经网络的更好的风险界。最近，在[25]中研究了将一个从其他任务训练的模型转移到当前任务的风险，只有一种特定的元学习方法。总之，到目前为止，只有在特定假设下的一些方法是有界的。

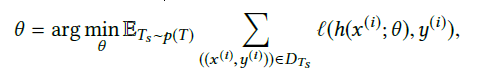
最后，算法中方法的收敛性还没有完全理解。事实上，用一些例子和先验知识来搜索h在并不能保证获得最小的经验风险。特别是对于元学习方法，他们优化了θ的任务分布，而不是单个任务的经验风险。目前还不清楚这种学习方式的效果。最近，在[37]中对此进行了分析。它们为一类元学习方法的收敛提供了充分的条件：元学习者是深层模型的底层，学习者是深层模型的最后一层，并通过梯度下降法进行优化。然而，对元学习方法收敛性的一般性或通用性分析仍然缺乏。

**7.总结**

很少有射击学习（FSL）的目标是弥合人工智能与类人学习之间的鸿沟。它可以通过融合先验知识来学习有限监督信息的新任务。FSL作为人工智能的试验台，有助于减轻为工业应用收集大规模监督数据的负担，或使学习罕见案例成为可能。随着人工智能的学术梦想和廉价学习的工业需求，FSL引起了人们的广泛关注，成为一个热门话题。在这项调查中，我们提供了一个全面和系统的检讨。本文首先对FSL进行了形式化定义，讨论了FSL与半监督学习、不平衡学习、迁移学习、元学习等相关学习问题的联系与区别。然后，指出了机器学习中基于误差分解的模糊学习的核心问题。我们发现，正是不可靠的经验风险最小化使得FSL难以学习。这可以通过满足或降低学习的样本复杂度来缓解。了解核心问题有助于根据如何利用先验知识解决核心问题，将不同的工作分为数据、模型和算法三类：数据增加了FSL的监督经验，模型约束了FSL的假设空间，算法改变了对假设空间中最优假设参数的搜索策略来求解FSL。在每一个范畴内，我们都会深入讨论不同范畴的利弊，并对每一个范畴下的见解进行总结。为今后的工作提供了可能的方向，包括问题设置、技术、应用和理论探索，以期对未来的FSL研究有所启发。

**A 元学习**

元学习者的学习需要大量的数据。设p（T）为任务T的分布。在量表学习中，它从一组任务Ts∼p（T）中学习。每个任务Ts操作N个类的数据集DTs，其中DTs={，}由训练集和测试集组成。每个学习者从中学习并测量上的测试错误。元学习者的参数θ通过



然后在元测试中，使用另一组不相交的任务∼p（T）来测试元学习者的泛化能力。每个Tt在N'类的数据集DTt上工作，其中={，}。学习者从训练集和测试集中学习。Tt的平均损失被视为元学习测试误差。为了理解元学习，图14展示了它的设置。

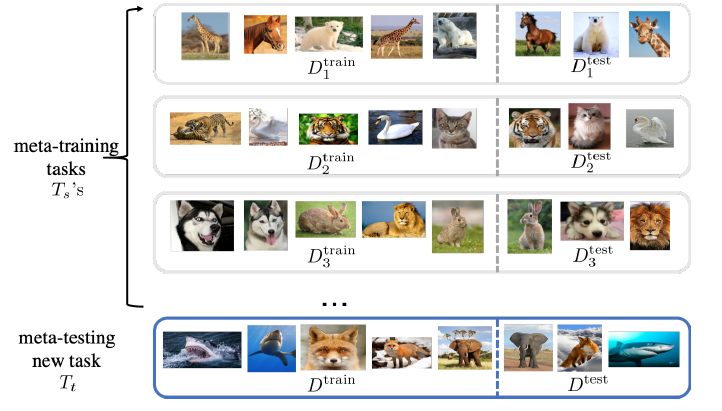


图14。元学习设置。该图改编自[93]。