TapNet: Neural Network Augmented with Task-Adaptive Projection for Few-Shot Learning

**会议：ICML2019**

**代码：**[**https://github.com/istarjun/TapNet**](https://github.com/istarjun/TapNet)

**个人看法：**这篇文章的参考向量集应该是随机初始化的，并在后续的训练中不断的使用opt进行更新。

**Abstract**

在仅仅给出几个训练示例之后，处理以前未见的任务仍然是机器学习中的一个艰巨挑战。我们提出了TapNets，神经网络与任务自适应投影相结合，以改进few-shot学习。在这里，使用基于事件的训练的元学习策略，一个网络和一组每类参考向量被学习到广泛变化的任务。同时，对于每一集，嵌入空间中的特征被线性地投影到一个新的空间中，作为快速任务特定条件反射的一种形式。训练损失是基于查询与投影空间中参考向量之间的距离度量得到的。这样的泛化效果很好。当在Omniglot、miniImageNet和tieredImageNet数据集上进行测试时，我们获得了各种few-shot场景下的最新分类精度。

**1. Introduction**

Few-shot学习承诺允许机器执行以前未见的任务，只使用少量相关的例子。因此，few-shot学习发现了广泛的应用，其中标记的数据是稀缺的或昂贵的，这是比没有数据更经常的情况。不幸的是，尽管近年来人们对机器学习产生了极大的兴趣并进行了积极的研究，但是few-shot学习对机器学习社区来说仍然是一个难以捉摸的挑战。例如，尽管deep networks现在通常在经过充分训练的标准图像测试数据集上提供近乎完美的分类分数，但few-shot学习的报告结果仍然远远低于在关键的现实环境中被认为可靠的水平。

开发few-shot学习策略的一种流行方法是采用元学习视角，结合情景训练（Vinyals等人，2016；Ravi&Larochelle，2017；Chen等人，2019）。元学习似乎向不同的人传达了不同的含义，但没有人会不同意它是关于学习新任务的一般策略（Vanschoren，2018）。情景训练是指一种训练方法，在这种方法中，各种各样的任务（或事件）被一个接一个地呈现给学习模型，每一个事件只包含几个有标记的例子。在最初的学习或元训练阶段，每次使用低样本重复暴露于以前未见的任务，似乎为学习者快速适应新数据提供了一个可行的选择（Vinyals等人，2016）。

在这方面的著名方法包括基于度量的学习者，如匹配网络（Vinyals等人，2016）和原型网络（Snell等人，2017）。这些方法都包含了非参数的、基于距离的学习，其中嵌入空间被训练为在稳定下来执行实际的few-shot分类之前最小化跨片段的相关距离度量。匹配网络训练独立的网络处理标记样本和查询样本，并利用嵌入空间中的每个标记样本作为参考点对查询样本进行分类。原型网络只使用一个嵌入网络，每个类的质心作为嵌入空间中的分类参考。基于只学习一个前馈特征抽取器，原型网络提供了一个令人惊讶的很好的能力，以推广到新的任务，因为一个归纳偏见似乎不知何故通过episodic training解决了。

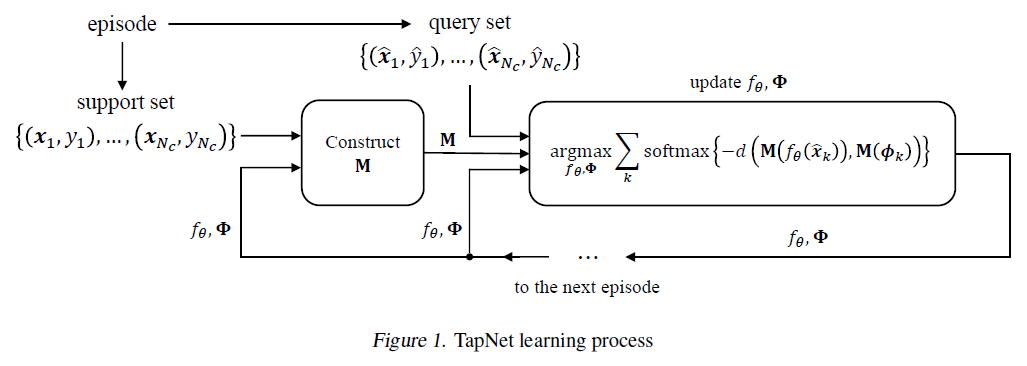
我们也对基于距离的学习感兴趣，除了阶段性元训练阶段外，没有参数微调。与以往的工作相比，该方法的独特之处在于通过嵌入特征的线性投影实现了显式任务相关条件反射。一旦神经网络的输出被投影到一个新的空间中，就可以根据每个类的参考向量之间的距离进行分类。神经网络和参考向量都是在反映广泛变化任务的事件序列中学习的，而投影的分类空间则是针对每个事件重新构造的。投影到一个任意分类空间是通过线性消除嵌入特征和每类引用之间的错误来实现的。与（Vinyals等人，2016）和（Snell等人，2017）不同，我们方案中类表示向量的类不是嵌入函数的输出。相反，我们的方法中的引用是一个简单的独立向量集，不直接耦合到输入图像，尽管它们是根据与投影在分类空间中的嵌入查询图像的距离为每个episode更新的。

将网络的跨任务学习和每类参考向量与分类空间的快速任务自适应调节相结合，可以实现出色的泛化。对Omniglot、minimagenet和tieredImageNet数据集进行的大量测试表明，采用任务自适应投影（TapNet）增强的网络可以获得最新的少量镜头分类精度。

**2. Task-Adaptive Projection Network**

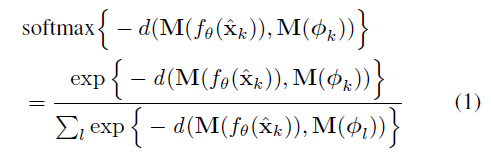
**2.1. Model Description**

TapNet由三个关键元素组成：嵌入网络、每类参考向量集、任务相关自适应投影或将嵌入特征映射到新的分类空间。是一个矩阵，其第k行是每类引用（行）向量。M表示投影或映射，但有时表示投影空间本身。参见图1，其中在连续的幕式训练过程中向模型呈现新的幕式。一集由一组支持图像/标签和一个查询集组成。为了清楚地说明这一点，对于每个给定的Nc类，在这里的任意一组中只有一个图像/标签对。



给定新的支持集，以及通过上一集阶段学习到的和，首先构造投影空间M，使得嵌入的特征向量和具有匹配标签的类参考向量在投影到M时紧密对齐。投影空间建设的细节将很快给出。

网络和参考集依次基于映射查询图像和参考向量之间的欧几里德距离d（）根据softmax进行更新：



在所有类k上取平均值。这里，M（z）表示行向量z的投影，并且除非另有规定，否则包括输入x的嵌入特征向量（x）的所有向量都假定为行向量。更新后的和被传递到下一集处理阶段。投影和参数更新继续为每一集，直到所有给定的集都用尽。

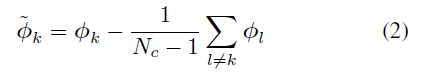
在少数镜头测试阶段，再次计算投影空间，将网络输出处出现的镜头的嵌入特征与参考点对齐，这两个参考点在经过整个情节元训练过程学习后现在都是固定的。最后将查询图像与投影空间中的参考图像进行比较，最终进行分类。

总之，嵌入者和每类参考向量是在不同任务（情节）之间学习的，而投影空间M是针对给定任务构建的，提供了快速的任务相关条件。由于大量的实验结果将很快得到验证，这种组合将产生一种极好的概括新数据的能力。

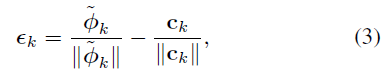
**2.2. Construction of Task-Adaptive Projection Space**

找到映射函数或投影空间M的基础是消除task-embedded特征和引用之间的不一致。若要处理每个类具有多个示例图像的一般情况，请将ck设为与支持集外的图像相对应的类k的嵌入特征的每个类平均值。

我们希望找到一个映射M，使得ck和匹配的参考向量在映射空间中高度对齐。同时，将所有的ck和非匹配权值在同一空间内很好地分离是有益的。结果发现，一个简单的线性投影，不需要任何学习，提供了一个有效的解决方案。其思想是找到一个投影空间，其中ck与一个修改过的向量对齐



式中，系数提供反映非匹配向量数量的自然归一化。也就是说，给定定义为



M使得每个k的投影误差向量为零，此时M是一个列数跨越投影空间的矩阵。换言之，M是通过误差的线性零化来找到的。被规范化以消除它们之间的功率不平衡。正式地，我们表达



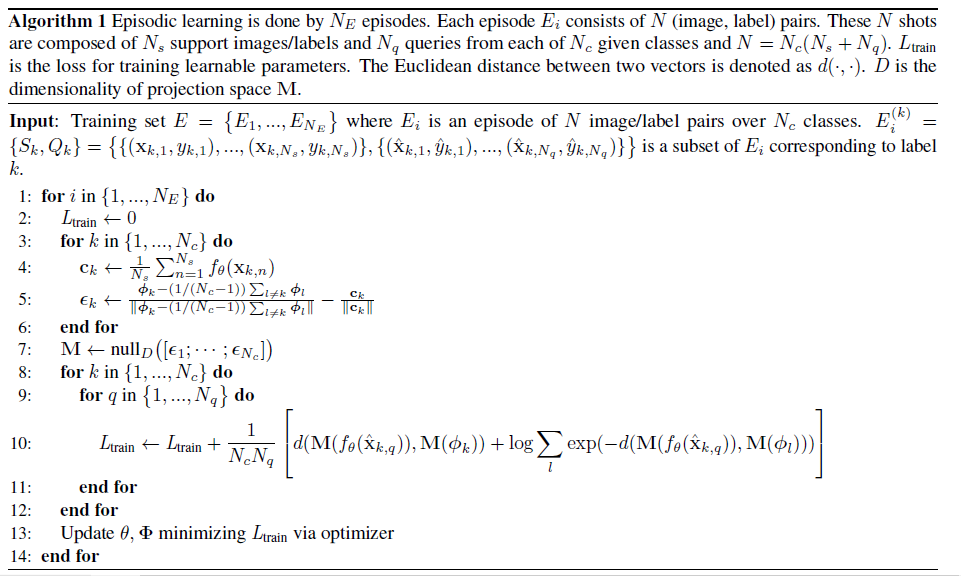
其中D是M的列维数。一个常用的方法是通过矩阵的奇异值分解（SVD），即取矩阵的右奇异向量D，从索引Nc+1到Nc+D，当L为的长度时，如果，则始终存在具有D维的投影空间M。

请注意，D可设置为小于，这表明有可能显著减小尺寸。我们的经验观察表明，降维有时实际上会提高很少的镜头分类精度。注意，对于具有的（nX m）矩阵的SVD，计算复杂度为O（mn2）。因此，获得投影所需的SVD计算复杂度为，与典型模型复杂度相比较小。

我们还注意到，由于如上所述的线性零陷的解决方案是存在的，而与的特定标记无关，因此我们不需要在每一个片段中重新标记参考向量；在整个片段训练阶段中，相同的标签粘在每个参考向量上。

**2.3. Training**

如前所述，嵌入网络和参考向量的训练通过幕式训练完成，如下（Vinyals等人，2016）。算法1给出了学习TapNets的具体步骤。



对于每个训练集，从给定数据集的训练集中随机选择Nc类。然后，对每一类，随机选取Ns标记的样本作为支持集Sk，Nq标记的样本作为查询集Qk，两个集之间不存在重叠样本。利用支持集Sk，得到每个类（第5行）的平均网络输出向量ck。基于每类平均网络输出向量，在不重新标记参考向量的情况下，获得所有类（第6行）的错误向量。然后，如前一小节所述，投影空间M被计算为误差信号的零空间。对于每个查询输入，测量到投影空间M中的参考向量的欧氏距离，并使用这些距离计算训练损失。在每个Nc给定类的所有Nq查询输入上获得平均训练损失（在第11行到第14行）。嵌入网络的可学习参数和参考参数现在基于平均训练损失进行更新（第16行）。这个过程会在剩下的每一集中使用新的图像和查询类重复。

接下来（Snell等人，2017年），我们还使用了比Nc更多的课程来进行阶段性训练，以提高绩效。例如，在幕式学习或元训练中使用20路分类，而在最后few-shot学习和测试中使用5路分类。在这种情况下，一个问题是如何在few-shot阶段线性投影的20个已训练参考向量中选择5个参考向量。为此，我们首先获得5个给定类的平均向量ck，然后选择最接近ck的5个参考向量，并在执行线性投影之前相应地重新标记它们。注意，使用更高的训练方式，TapNets可以轻松处理实际很few-shot分类的类数事先未知的情况。