

TrAdaBoost

起初给训练数据 T 中的每一个样例都赋予一个权重，当 target domain 中的样本被错误的分类之后，我们认为这个样本是很难分类的，于是乎可以加大这个样本的权重。如果 source domain 中的一个样本被错误的分类了，我们认为这个样本对于目标数据是很不同的，我们就降低这个数据在样本中所占的权重。

迁移成分分析:

类似于 PCA，都是非监督的学习方法，PCA 用于降维，TCA 是为了发现共同潜在特征，将 source 和 target domain 都映射到一个空间，取某些维度的数据，这些数据就是共同潜在特征。不仅有相同的分布，同时还保留有初始域数据的内在结构。这种潜在特征通过将源域和目标域映射到再生希尔伯特核空间学习得到，使用最大平均差异作为边缘分布度量标准。一旦潜在特征被发现，就可以使用传统的机器学习方法训练最终目标分类器。

MMD:

最大均值差异选择一些常用的核函数进行映射（比如线性核、高斯核），然后求映射后源域和目标域的均值之差。

MK-MMD:

在 MMD 中这个核函数是固定的，我们在实现的时候可以选择是高斯核还是线性核。

MK-MMD 提出用多个核去构造这个总的核，对多个 kernel 加权，形成一个总的核。

DAN:

在 DAN 中，所有任务特定层的隐藏表示都嵌入到一个再生 Hilbert 空间中，在这里可以将不同域数据表示嵌入到同一空间中。采用一种最优多核选择方法进行平均嵌入匹配，进一步减小了域差异。

源域和目标域的数据都放在一起，通过 AlexNet 来训练，前三层 frozen，学习率为 0，第四层第五层 fine-tuning，当到后面几层时，source data 和 target data 分开，然后通过 MK-MMD 方法来计算两个域的距离，并且通过损失函数来进行优化，最后当损失函数优化到设定的阈值时，就可进行最终的分类

CoVe:

在计算机视觉领域中，使用预先训练的特征表示“微调”预训练模型比较普遍。常用的网络有 LeNet, AlexNet, VGG, Inception, ResNet

那么在 NLP 中，也可以把一个任务中训练好的模型用于另外一个任务。基于这样的想法，作者提出了将 context vectors (CoVe) 添加到原有的模型中的方法，并且通过实验证明在常见的 NLP 任务，例如情感分析、问题分类、推理和问答等，都起到了提升性能的作用。

a 是一个典型的 encoder-decoder 模型，经常被用于机器翻译中。在这个工作中，作者训练了一个两层双向的 LSTM 模型作为基于注意力的 sequence-to-sequence 机器翻译模型的 encoder 并且将其应用于其他的 NLP 任务。具体的，作者通过英语-德语机器翻译任务训练了一个神经网络模型（应该就是 ab 中的 Encoder 部分），然后得到上下文向量——CoVe。然后将 CoVe 和常用词向量通过一定方式结合得到用于特定任务的输入，最终提升任务模型的性能。

unsupervised domain adaptation by backpropagation:

是结合对抗网络框架与深度学习技术解决 domain adaptation 应用的一个工作。具体而言，在这个框架中对三个部分进行训练：一个是 feature extractor，这个是为了提取特征的，一般由卷积层与 pooling 层组成；另一个是 label classifier，预测分类标签（训练与测试的时候），使用全连接层+逻辑斯蒂分类器；第三个是构成对抗网络框架的分类器 domain classifier，用于区分数据来自 source domain 还是 target domain，由全连接层+交叉熵分类器构成。

对抗体现在对于 domain classifier 损失在训练阶段两个相反的要求。具体而言：对于预测分类标签来说，我们希望网络学到的特征表示具有领域不变（domain invariant）的特征，那么就要求 domain classifier 不能正确进行领域分类，也就是要求 domain classifier 的分类损失最大，这样学到的特征尽可能具有 label 的分类判别信息，也就是最小化 label classifier 的分类损失；另一方面在对 domain classifier 训练时，我们肯定要求分类器能尽可能的正确分类出 source 和 target domain 的区别，也就是 domain classifier 的分类损失最小。