

一种基于伪标签的半监督少样本学习模型

余 游,冯 林,王格格,徐其凤
(四川师范大学计算机科学学院,四川成都 610101)

摘 要: 如何将带有大量标记数据的源域知识模型迁移至带有少量标记数据的目标域是少样本学习研究领域的热点问题. 针对现有的少样本学习算法在源域数据与目标域数据的特征分布差异较大时存在的泛化能力较弱的问题,提出一种基于伪标签的半监督少样本学习模型 FSLSS(Few-Shot Learning based on Semi-Supervised). 首先,利用 pytorch 深度学习框架建立一个关系型深度学习网络,并使用源域数据对网络进行预训练;然后,使用此网络对目标域数据进行分类预测,将分类概率最大的类标签作为数据的伪标签;最后,利用目标域的伪标签数据和源域的真实标签数据对网络进行混合训练,并重复伪标签标记与混合训练过程. 实验结果表明,相对于现有主流少样本学习算法, FSLSS 模型有更好的泛化能力及知识迁移效果.

关键词: 少样本学习; 半监督学习; 伪标签; 迁移学习

中图分类号: TP181

文献标识码: A

文章编号: 0372-2112 (2019)11-2284-08

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.3969/j.issn.0372-2112.2019.11.007

A Few-Shot Learning Model Based on Semi-Supervised with Pseudo Label

YU You, FENG Lin, WANG Ge-ge, XU Qi-feng

(Department of Computer Science, Sichuan Normal University, Chengdu, Sichuan 610101, China)

Abstract: How to migrate a source domain knowledge model with a large amount of tagged data to a target domain with a small amount of tagged data is a hot issue in few-shot learning. For the problems that the existing few-shot learning algorithm have weak generalization ability when the difference between the feature distribution of the source domain data and the target domain data is large, a few-shot learning model based on semi-supervised FSLSS(Few-Shot Learning based on Semi-Supervised) is proposed. Firstly, a relational deep learning network is established by using the pytorch framework, and the network is pre-trained by the source domain data. Then, the network is used to predict the target domain data, and the label with the highest classification probability is used as the data's pseudo label. Finally, the network is hybrid trained using the pseudo label data of the target domain and the real label data of the source domain, then repeating the pseudo-labeled and hybrid trained process. The experimental results show that the FSLSS model has better generalization ability and knowledge transfer effect than the existing few-shot learning algorithms.

Key words: few-shot learning; semi-supervised learning; pseudo label; transfer learning

1 引言

自深度学习理论与方法提出以来,它已在机器学习、模式识别等研究领域引起了学术界与产业界广泛关注,大量的研究成果在 Nature、Science、计算机学报、电子学报等权威期刊发表^[1-5]. 但是,深度学习模型存在两个主要不足:(1)需要数量庞大的带标记数据训练模型;(2)一旦训练模型的数据发生改变,则需花费较大代价重新训练模型. 迁移学习为解决这些问题提供

了理论与方法支撑^[6]. 迁移学习是指在源域数据上建立的知识模型,可迁移至与源域具有一定差别的目标域上使用且能达到较高的识别率. 作为迁移学习研究领域的一个分支,少样本学习已经成为当前机器学习领域的研究热点^[7-16]. 较早的研究成果中,少样本学习大多基于统计学习方法. 例如, Maas 等^[9]提出的贝叶斯网络,其方法是使用概率分布捕获属性之间的关系,此方法允许两种关系,即属性值是其父属性值的软概率关系,以及在给定父元素值的情况下以近似确定的方

收稿日期:2019-05-06;修回日期:2019-07-01;责任编辑:马兰英

基金项目:国家科技支撑计划项目(No. 2014BAH11F01, No. 2014BAH11F02);四川师范大学 2019 年研究生优秀论文培育基金(No. 川师研 201903-36)

式生成属性值的关系. 随着神经网络的发展, 现有的主流少样本学习算法大多基于神经网络. 例如: Koch 等^[11]提出了孪生网络 (Siamese Neural Networks), 其主要思想是: 一次将两张图片输入到两个完全相同的深度网络 (权重共享), 然后将两个网络输出的绝对差值送到一个线性分类器, 判断这两张图片是否是一类.

由于在少样本学习中, 源域与目标域的类别空间不相交, 并且目标域仅提供少量标记数据, 因此如何提升少样本学习分类模型的泛化能力是迁移学习中一个很富有挑战性的问题. 现有少样本学习模型存在的缺陷是, 它们在源域数据上训练出来的模型, 直接应用在目标域数据上, 没有考虑到源域数据与目标域数据之间的特征分布差异, 导致训练出来的模型分类准确率不高. 因此, 需要从理论上解决在源域数据和目标域数据特征分布差异很大的情况下, 现有少样本分类模型的泛化能力弱的问题. 另一方面, 半监督学习算法作为机器学习的新型分支, 其不仅能解决监督学习模型泛化能力弱的问题, 而且能解决无监督学习模型不精准的问题, 进而为提升少样本分类模型的泛化能力提供了新思路.

本文针对上述问题, 提出了一种基于伪标签的半监督少样本学习算法 FSLSS. 首先, 利用 pytorch 深度学习框架建立一个关系型深度学习网络, 并使用源域数据对网络进行预训练; 然后, 使用此网络对目标域数据进行分类预测, 将分类概率最大的类标签作为数据的伪标签; 最后, 利用目标域的伪标签数据和源域的真实标签数据对网络进行混合训练, 并重复伪标签标记与混合训练过程. FSLSS 模型充分考虑了目标域数据的分布情况, 并用以指导模型的训练, 实现了提升模型的泛化能力.

2 理论基础

2.1 少样本学习的概念

目前的研究成果中, 少样本学习的相关概念均基于文字描述, 缺少相应的数学形式化定义. 为了方便叙述, 本文基于文献[11~16]的论述, 首先用数学定义形式化与少样本学习相关的基本概念.

定义 1 (数据集) 一个数据集 D 是一个三元组 $D = (X, Y, f)$, 其中:

$X = \{x_1, x_2, \dots, x_n\}$ 是 n 个输入实例观察值集合, 本文中, $\forall x_i \in X, x_i$ 表示输入的一个图片实例.

$Y = \{y_1, y_2, \dots, y_m\}$ 是 m 个类别标签集合.

$f: X \rightarrow Y$ 是一个信息函数, 它指定 X 中每一个输入实例的类别标签值, 即对 $\forall x_i \in X, \exists y_j \in Y$, 有 $f(x_i) = y_j$ 成立 ($i = 1, 2, \dots, n; j = 1, 2, \dots, m$).

定义 2 (类标签实例集) 给定数据集 $D = (X, Y, f)$, $\forall y_j \in Y$, 若 $f^{-1}(X|y_j) = \{x_i \in X | f(x_i) = y_j\}$, 则称 $f^{-1}(X|y_j)$ 为数据集 D 下的类别标签 y_j 的实例集 ($i = 1, 2,$

$\dots, n; j = 1, 2, \dots, m$).

定义 3 (训练集、支持集与测试集) 给定数据集 $D = (X, Y, f)$, 以及参数 C, K , 随机生成 D 的三个子集 $T_s = (X_s, Y_s, f), S_u = (X_u, Y_u, f), T_e = (X_e, Y_e, f)$, 如果满足:

- ① $Y_u = Y_e$, 且 $Y_u \cap Y_s = \emptyset, Y_u \cup Y_s = Y$.
- ② $X_u \cap X_e = \emptyset$.
- ③ $|Y_u| = C$, 其中 $|\bullet|$ 表示集合的势.
- ④ $\forall y_i, y_j \in Y_u, |f^{-1}(X_u|y_i)| = |f^{-1}(X_u|y_j)| = K$.

则称集合 T_s, S_u, T_e 分别为 D 上的训练集、支持集与测试集.

定义 4 (样本集与查询集) 给定训练集 $T_s = (X_s, Y_s, f)$, 以及参数 C, K , 设 $S_a = (X_a, Y_a, f)$ 和 $Q_r = (X_r, Y_r, f)$ 为 T_s 的两个子集, 如果满足:

- ① $Y_a = Y_r, |Y_a| = |Y_r| = C$.
- ② $\forall y_i, y_j \in Y_a, |f^{-1}(y_i|X_a)| = |f^{-1}(y_j|X_a)| = K$.
- ③ $X_a \cap X_r = \emptyset$.
- ④ $|X_a| + |X_r| = \sum_{y_i \in Y_s} |f^{-1}(X_s|y_i)|$.

则称 S_a, Q_r 为训练集 T_s 上的样本集和查询集.

定义 5 (C-way, K-shot 少样本学习) 给定数据集 $D = (X, Y, f)$, 及参数 C, K , 根据定义 3, 划分出训练集 T_s 、支持集 S_u 、测试集 T_e . 给定任务: 利用训练集 T_s 训练出一个分类模型, 它能在支持集 S_u 的支持下较好应用于测试集 T_e . 称这个任务为 C -way, K -shot 少样本学习.

2.2 半监督学习模型

半监督学习 (Semi-Supervised Learning, SSL)^[17] 是介于监督学习和无监督学习之间的一种新型机器学习算法, 如图 1 所示, 其主要思想是在少量标记样本下, 通过引入大量的未标记样本辅助模型训练, 以避免模型在训练集上出现过拟合等情况, SSL 不仅能解决监督学习模型泛化能力弱的问题, 而且能解决无监督学习模型不精准的问题. 半监督学习的实现^[18] 需要基于三个假设: 平滑假设、聚类假设、流形假设, 大量实验研究表

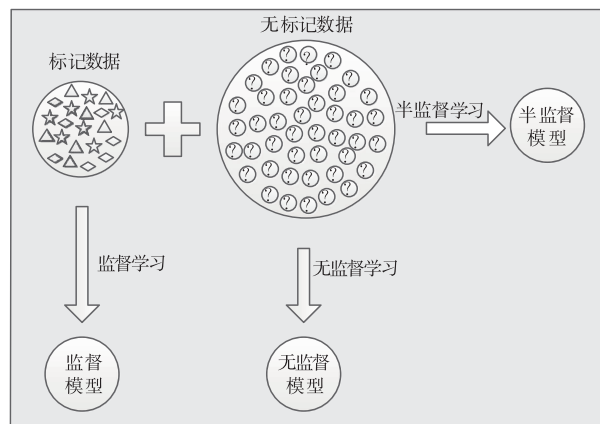


图1 半监督学习

明^[19,20],当 SSL 中不满足以上假设时,引入无标记数据不仅不会改善模型的学习性能,还有可能对模型起到负面影响.

2.3 卷积神经网络

2006 年,Hinton 团队^[21]提出了深度置信网络的训练方式,随着十几年的发展,深度学习网络已经在图像识别^[22]、语音识别^[23]、文本处理^[24]等领域中取得巨大的成功.卷积神经网络(Convolutional Neural Networks, CNNs)^[25]是深度网络中经典的学习方法,其中包括卷积层、池化层、激活函数、损失函数等重要的概念.

卷积层和池化层:卷积层由若干个卷积核组成,通过对输入矩阵进行卷积操作,可提取出输入矩阵的不同特征.池化层也称采样层,常用的有平均池化层和最大池化层,其主要目的是对输入特征进行压缩映射,提取主要特征.

激活函数:用来加入非线性因素的,提高神经网络对模型的表达能力,常用的激活函数有 Sigmoid 函数、Tanh 函数、Relu 函数、Softmax 函数等.本文拟采用 Relu 函数和 Sigmoid 函数.

损失函数:用来评估模型预测值 Y' 与真实值 Y 之间的差异程度.网络的训练过程,是通过梯度下降法回传损失函数的值,然后相应调整模型的参数,使损失函数达到最小的过程.常用的损失函数是均方误差损失函数和交叉熵损失函数.本文拟采用均方误差函数.

3 FSLSS 模型

FSLSS 模型的流程如图 2 所示,它分成关系型网络、预训练、半监督训练三个部分.本文先对三个部分的内容进行介绍.然后,再对 FSLSS 模型进行详细描述.

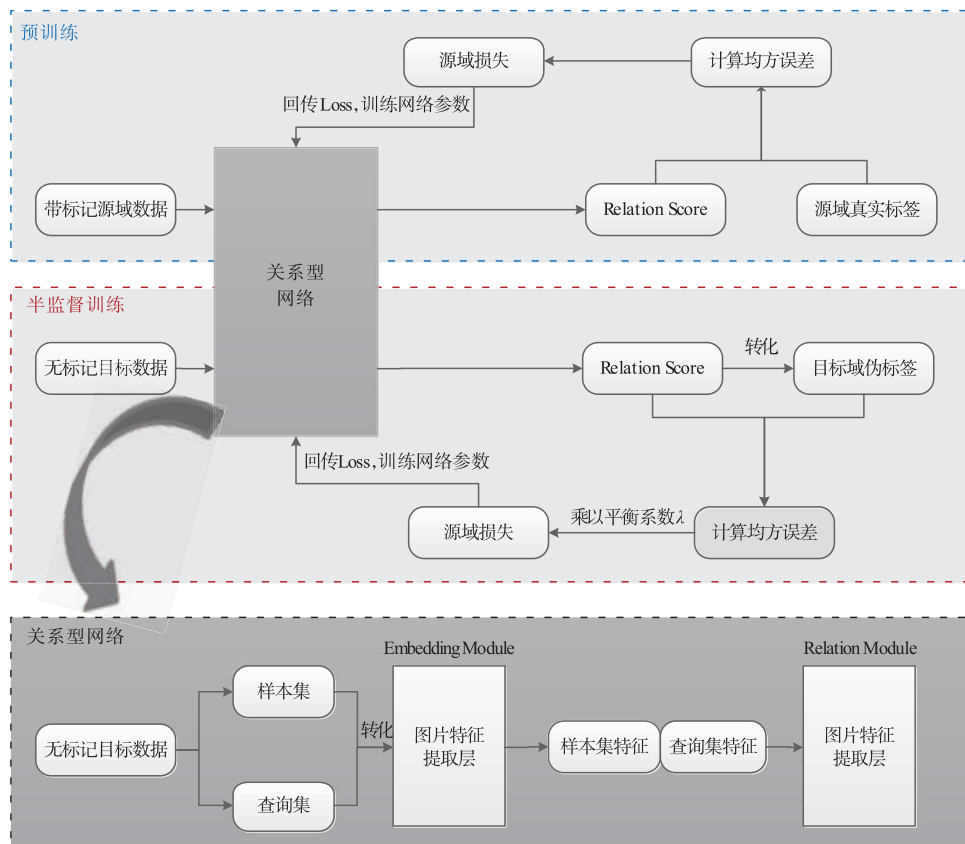


图2 FSLSS模型描述

3.1 关系型网络

关系型网络(Relation Network)是一个卷积神经网络学习器,其网络结构及其参数配置如表 1 所示.其中包含两个模块:Embedding Module、Relation Module. Embedding Module 由 4 个卷积层和 2 个最大池化层组成,主要用于提取图像特征;Relation Module 由 2 个卷积层、2 个最大池化层、2 个全连接层组成,其主要是用于

图片之间的相似度回归.整个 Relation Network 中,所有的卷积核大小为 3×3 、池化核大小为 2×2 .关系型网络的核心思想是:首先,根据定义 4,取出样本集 S_a 与查询集 Q_r ,并利用 Embedding Module 分别提取出 S_a 与 Q_r 的图像特征 ξ_1, ξ_2 ;其次,利用 relation 网络计算出 ξ_1, ξ_2 之间的相似度,其值代表着每个样本集的图片实例与每个查询集的图片实例的相似程度,取出与查询集

图片实例相似度最高的类别标签作为当前样本集的图片实例的预测标签。

表 1 关系型网络结构

Relation Network		
Layer	Embedding Module	Relation Module
1	Conv (Num = 64, Kernel = 3 × 3, Pad = 0, name = H ₁), BN, ReLU	Conv (Num = 64 * 2, Kernel = 3 × 3, Pad = 0, name = H ₇), BN, ReLU
2	Max-pool (Kernel = 2 × 2, name = H ₂)	Max-pool (Kernel = 2 × 2, name = H ₈)
3	Conv (Num = 64, Kernel = 3 × 3, Pad = 0, name = H ₃), BN, ReLU	Conv (Num = 64, Kernel = 3 × 3, Pad = 0, name = H ₉), BN, ReLU
4	Max-pool (Kernel = 2 × 2, name = H ₄)	Max-pool (Kernel = 2 × 2, name = H ₁₀)
5	Conv (Num = 64, Kernel = 3 × 3, Pad = 1, name = H ₅), BN, ReLU	Linear (name = W ₁)
6	Conv (Num = 64, Kernel = 3 × 3, Pad = 1, name = H ₆), BN, ReLU	Linear (name = W ₂)

3.2 预训练

在开始半监督训练前,需要对网络模型进行预训练.先利用 pytorch 深度学习框架和表 1 的网络结构建立一个关系型深度学习网络.其次,将源域数据集划分成训练集、验证集、测试集,并将训练集输入到关系型网络中.然后,利用式(1)计算出关系型网络的输出 Relation Score 与训练集的真实标签的均方误差值.最后,将均方误差值作为网络的损失,回传给网络,并利用梯度下降法训练网络参数.

$$L_r(Y, F) = \frac{1}{n} \sum_{m=1}^n \sum_{i=1}^C (y_i^m - f_i^m)^2 \quad (1)$$

其中: n 代表样本个数, C 代表样本种类, F 代表网络的输出值 Relation Score, Y 代表样本的真实标签.

3.3 半监督训练

随着十几年的发展,半监督学习算法已经成熟,自训练(Self-Training)或伪标签(Pseudo-Label)半监督学习也成为一种很好的半监督学习方式^[26].伪标签半监督训练过程是:首先,将目标域无标记数据输入到预训练后的关系型网络模型中;其次,使用预训练的模型对无标记目标域数据进行预测,即为无标记目标域数据标记上伪标签.然后,利用式(2)计算出目标域数据的半监督损失,回传给网络,并利用梯度下降法继续训练网络参数.最后,将带标记源域数据和新生成的伪标签目标域数据作为新的训练数据对网络模型进行混合

训练.

$$L_p(Y', F) = \lambda \frac{1}{n'} \sum_{m=1}^{n'} \sum_{i=1}^C (y_i^m - f_i^m)^2 \quad (2)$$

其中: n' 代表样本个数, C 代表样本种类, F 代表网络的输出值 Relation Score, Y' 代表样本的真实标签, λ 是一个平衡系数(一般取值在 0~1 之间),其值的大小决定了无标记数据对模型的影响程度,其值越大代表引入的无标记数据对模型的影响程度越大.

3.4 FSLSS 模型的实现

根据定义 5,可知,少样本学习算法是在两个类标签互不相交的域中进行迁移学习.本文,将定义 3 中的带标记的训练集视为半监督学习算法中的带标签源域数据、将带少量标记数据的测试集视为半监督学习算法中的无标签目标域数据,于是将少样本学习算法转化成一种特殊的半监督学习.基于此,提出一种基于伪标签的半监督少样本学习模型 FSLSS,其算法描述如下:

算法 1 FSLSS 学习策略

Input: 数据集 $D = (X, Y, f)$ 以及参数 $C, K, \lambda, Times, \theta$; // C 代表分类种类, K 代表每个种类的样本数, λ 是平衡系数, $Times$ 代表最大训练次数, θ 代表损失函数收敛的一个阈值.

Output: 少样本分类模型 Ω .

Step1 根据定义 3, 从数据集中 D 中划分出数据集 T_s, S_n 和 T_c ;

Step2 随机初始化模型 Ω 中所有网络层的参数;

Step3 while ($Times > 0$ and $L_p < \theta$):

3.1. 取出训练集 T_s , 根据定义 4, 随机划分出 $S_n = (X_n, Y_n, f)$ 和 $Q_r = (X_r, Y_r, f)$;

3.2. 取出 S_n 中所有图片实例集 X_n , 按照步骤 3.2. 的方法计算 X_n 的图像特征, 记为 ξ_1 ;

3.3. 取出 Q_r 中所有图片实例集 X_r , 按照步骤 3.2. 的方法计算 X_r 的图像特征, 记为 ξ_2 ;

3.4. 将 ξ_1, ξ_2 拼接成一个联合特征矩阵 ξ ;

3.5. η 是一个高维矩阵, 其每一行代表着样本集中的一个样本实例与查询集中每个类别的相似度 $F = \{f_1, f_2, \dots, f_n\}$;

3.6. 将支持集 S_n 视为样本集 S_n 、测试集 T_c 视为查询集 Q_r , 按照 3.2. — 3.4. 的步骤, 计算出测试集的 η' , 其每一行中的每一个元素, 代表着支持集中的一个样本实例与测试集中一个类别的相似度 $F = \{f_1, f_2, \dots, f_n\}$;

3.7. 取 η' 中每行最大相似度所对应的类别为该样本图片实例的预测标签, 记预测标签为 $Y' = \{y_1, y_2, \dots, y_n\}$;

3.8. 利用式(1)和式(2)计算模型 Ω 的监督损失 L_r 和 L_p ;

3.9. 利用梯度下降法反向传播 $L_r + L_p$ 的值, 以调整模型 Ω 中的参数;

3.10. $Times = Times - 1$;

Step4 输出少样本分类模型 Ω , 算法停止.

4 实验及结果分析

为了验证本文:模型的效果,实验分成三个部分:(1)与现有主流少样本学习算法的对比实验;(2)分析

平衡系数 λ 对 FSLSS 分类效果的影响; (3) 迁移效果分析.

实验数据集的选择: miniImageNet 和 Omniglot 是研究少样本学习算法的两个经典的公开数据集. miniImageNet^[11] 数据集含有 100 个常见物种, 每个物种之间的差异性很大, 例如麻雀和卡车, 每个物种有 600 张图片, 总共 60000 张. Omniglot^[27] 数据集包含来自 50 个不同字母的 1623 个不同手写字符. 每一个字符都是由 20 个不同的人通过亚马逊的 Mechanical Turk 在线绘制的.

本文实验采用的硬件环境为 NVIDIA Tesla K80 GPU 平台; 软件环境为 Linux 系统、Python 编程语言、Pytorch 深度学习框架.

4.1 评价准则

4.1.1 识别率

定义 6 (识别率) 给定测试集 $T_e = (X_e, Y_e, f)$ 和分类模型 Ω , 利用模型 Ω 预测出测试集的标签值 Y_e' , 利用公式(3)计算出 $acc(Y_e, Y_e')$, 其值称为模型 Ω 在数据集 $T_e = (X_e, Y_e, f)$ 的识别率. 一般来说, 模型的识别率越高, 模型越好.

$$acc(Y_e, Y_e') = \frac{1}{|Y_e|} \sum_{i=1}^{|Y_e|} Y_e(i) == Y_e'(i) \quad (3)$$

其中: $x == y$ 表示: 当 $x = y$ 时, 其值取 1; 当 $x \neq y$ 时, 其值取 0.

4.1.2 MMD 距离

MMD (Maximum Mean Discrepancy) 距离最先是由 Gretton 等人^[28] 提出的, 用于解决双样本的检测 (two-sample test) 问题, 其目的是判断两个分布 p 和 q 是否相同. 在迁移学习环境下, 一般使用 MMD 来度量两个分布的匹配程度, 其值代表着两个分布在再生希尔伯特空间 (带有内积的完备向量空间, RKHS) 中的距离, 其距离越近, 代表两个特征分布越相似. MMD 的计算公式如下:

$$\begin{aligned} MMD(X, Y) = & \left\| \frac{1}{n^2} \sum_i^n \sum_{i'}^n k(x_i, x_{i'}) \right. \\ & - \frac{2}{nm} \sum_i^n \sum_j^m k(x_i, y_j) \\ & \left. - \frac{1}{m^2} \sum_j^m \sum_{j'}^m k(y_j, y_{j'}) \right\|_H \quad (4) \end{aligned}$$

其中: H 表示这个距离是将数据映射到再生希尔伯特空间中进行度量的, $k(*)$ 表示一个核函数, 常采用如下高斯核函数:

$$k(u, v) = e^{-\frac{\|u-v\|^2}{\sigma}} \quad (5)$$

4.2 实验 1: 与现有少样本学习算法的对比实验

实验 1 的主要目的是测试 FSLSS 模型的有效性. 因此, 给出了 FSLSS 模型与目前主流的少样本学习方法的实验对比. 实验参数设置: $\lambda = 0.2$, Times = 50000, $\theta =$

0.0001, 学习率 = 0.001.

实验步骤如下.

Step1 按照定义 3, 将 miniImageNet 数据集划分成训练集 (64 类)、验证集 (16 类)、支持集与测试集 (20 类), 然后将训练集输出少样本分类模型.

Step2 分别取 ($C = 5, K = 1$) 和 ($C = 5, K = 5$), 按照算法 1 的步骤, 分别训练出少样本分类模型 Ω_1 和模型 Ω_2 .

Step3 利用 miniImageNet 的支持集与测试集随机产生多组子支持集与子测试集, 计算出 Ω_1 在每组子支持集与子测试集中的识别率, 然后取其平均值作为 Ω_1 的识别率. 同样的步骤, 计算出 Ω_2 在 miniImageNet 上的识别率.

Step4 按照定义 3, 取出 Omniglot 数据集的前 1200 类为训练集, 后 423 类为支持集与测试集, 将训练集输入到少样本分类模型.

Step5 分别取 ($C = 5, K = 1$)、($C = 5, K = 5$)、($C = 20, K = 1$) 和 ($C = 20, K = 5$), 按照算法 1 的步骤, 分别训练出分类模型 Ω_3 、 Ω_4 、 Ω_5 、 Ω_6 .

Step6 利用 Omniglot 的支持集与测试集随机产生多组子支持集与子测试集, 计算出 Ω_3 在每组子支持集与子测试集中的识别率, 然后取其平均值作为 Ω_3 的识别率. 同样的步骤, 计算出 Ω_4 — Ω_6 在 Omniglot 上的识别率.

4.2.1 miniImageNet 数据集实验结果

将 FSLSS 模型在 miniImageNet 数据集的识别率与 Matching Nets、Prototypical Nets、Model-Agnostic 和 Relation Nets 做实验对比, 实验结果如表 2 所示. 从表 2 中可以看出, 5 way 1 shot 设定下, FSLSS 模型的识别率达到了 $53.85 \pm 0.81\%$, 比 Relation Nets ($50.44 \pm 0.82\%$) 的识别率高出了 3.41%, 且超过了现有主流的其他少样本算法; 5 way 5 shot 设定下, FSLSS 模型的识别率达到了 $71.50 \pm 0.69\%$, 比 Relation Nets ($65.32 \pm 0.70\%$) 的识别率高出了 6.18%, 也均超过了其它主流的少样本学习算法.

表 2 miniImageNet 数据集下 FSLSS 模型与目前主流少样本学习方法识别率对比

Model	5way 识别率	
	1shot(Ω_1)	5shot(Ω_2)
Matching Nets ^[12]	43.56 \pm 0.84%	55.31 \pm 0.73%
Meta-Learn ^[15]	43.44 \pm 0.77%	60.60 \pm 0.73%
Prototypical Nets ^[8]	49.42 \pm 0.78%	68.20 \pm 0.66%
Model-Agnostic ^[16]	48.70 \pm 1.84%	63.10 \pm 0.92%
Relation Nets ^[13]	50.44 \pm 0.82%	65.32 \pm 0.70%
本文 FSLSS	53.85 \pm 0.81%	71.50 \pm 0.69%

其中:识别率 $A \pm B\%$ 中, A 代表多次测试的识别率的均值, B 代表其方差.

4.2.2 Omniglot 数据集实验结果

同样地, 将 FSLSS 模型在 Omniglot 数据集的识别率与 Matching Nets、Prototypical Nets、Model-Agnostic 和 Relation Nets 做实验对比, 实验结果如表 3 所示. 从表 3

表 3 Omniglot 数据集下 FSLSS 模型与目前主流少样本学习方法识别率对比

Model	5 way 识别率		20 way 识别率	
	1 shot (Ω_3)	5 shot (Ω_4)	1 shot (Ω_5)	5 shot (Ω_6)
Matching Nets ^[12]	98.1%	98.9%	93.8%	98.5%
Prototypical Nets ^[8]	98.8%	99.7%	96.0%	98.9%
Model-Agnostic ^[16]	98.7 \pm 0.4%	99.9 \pm 0.1%	95.8 \pm 0.3%	98.9 \pm 0.2%
Siamese Nets ^[11]	98.4%	99.6%	95.0%	98.6%
Relation Nets ^[13]	99.6 \pm 0.2%	99.8 \pm 0.1%	97.6 \pm 0.2%	99.1 \pm 0.1%
本文 FSLSS	99.8 \pm 0.2%	99.9 \pm 0.1%	99.1 \pm 0.2%	99.6 \pm 0.1%

4.3 实验 2: 分析平衡系数 λ 对 FSLSS 分类效果的影响

实验 2 的主要目的是为了研究平衡系数 λ 对 FSLSS 分类效果的影响. 从表 4 中可以看出在 Omniglot 数据集下, FSLSS 的准确度提升空间太小. 因此, 我们采用 miniImageNet 数据集来研究平衡系数 λ 对 FSLSS 分类效果的影响, 将 λ 依次取值 0、0.05、0.1、0.15、0.2、0.25、0.3, 分别在 5 way 1 shot 和 5 way 5 shot 参数设定下, 按照实验 1 的步骤进行对比实验. 实验步骤如下:

Step1 For i in [0, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3]

1.1. 令 $\lambda = i$, 按照实验一的步骤, 输出 Ω_1 的识别率和 Ω_2 的识别率;

Step2 记录每一次的 Ω_1 的识别率和 Ω_2 的识别率.

实验结果如图 3 所示, 从图 3 中可以看出, 随着平衡系数 λ 的增加, FSLSS 的分类准确率先上升后下降. 在 5 way 5 shot 设定下, λ 的值等于 0.2 的时, FSLSS 的分类精度达到最高, 超过 0.25 时, 模型的准确度迅速下降. 在 5 way 1 shot 设定下, λ 的值等于 0.15 的时, FSLSS 的分类精度达到最高, 超过 0.2 时, 模型的准确度迅速下降. 实验结果说明了平衡系数 λ 对 FSLSS 分类效果有很大的影响, λ 值太小会达不到最好的提升效果, λ 值太大会对模型优化造成负面影响. 所以, 在 FSLSS 模型中, 选取一个合适的 λ 值至关重要. λ 是经验值, 本文通过实验验证了在 FSLSS 模型中将 λ 的值设在 0.2 左右能达到不错的效果. 有的研究^[26]还将平衡系数 λ 设为变量, 模型刚开始训练时, 伪标签的准确率很低, 因此将 λ 设为 0, 随着模型准确率的上升, 慢慢将 λ 的值增大.

4.4 实验 3: 迁移效果分析

实验 3 的目的是分析 FSLSS 模型的迁移效果. 本文

中可以看出, 对于 Omniglot 数据集, 在 5 way 1 shot、5 way 5 shot、20 way 1 shot、20 way 5 shot 参数设定下, 本文 FSLSS 模型的识别率均超过了其它主流的少样本算法的识别率. 尤其是在具有挑战的 20 way 5 shot 参数设定下, FSLSS 模型的识别率都可达 $99.1 \pm 0.2\%$.

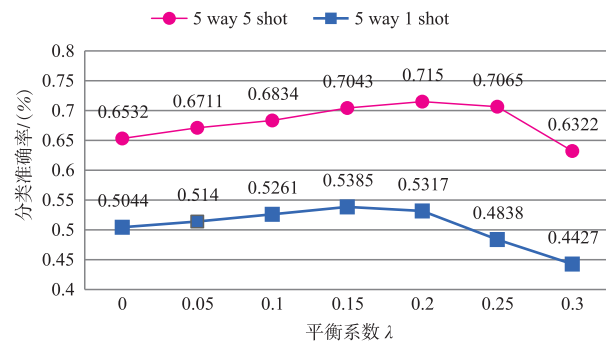


图 3 平衡系数 λ 对 FSLSS 分类效果的影响

采用 MMD 距离, 来度量 FSLSS 模型的迁移效果. 其实验步骤如下.

Step1 分别从源域数据和目标域数据中随机抽取 5 个类别, 每个类别里随机取 30 个样本数据, 共 150 个样本, 分别记为 X_1, X_2 .

Step2 移除网络模型 Ω 最后两层全连接层, 以获得网络提取的图片特征分布.

Step3 将 X_1, X_2 输入到移除掉全连接层的网络模型 Ω 中, 获得源域数据和目标域数据的特征分布, 记为 Z_1, Z_2 .

Step4 利用公式 4, 计算出 Z_1, Z_2 的 MMD 距离.

Step5 重复 Step1-Step4, 随机测试 4 次.

实验结果如表 4 所示, 其中: 未训练代表使用随机初始化网络参数的模型, 按照实验 3 的步骤测出源域数据和目标域数据的特征分布的 MMD 距离; 预训练代表使用预训练后的网络模型进行实验; 半监督训练代表使用 FSLSS 方法训练后的模型进行实验. 从表 4 中可以看出, 在刚开始没有训练时, 两个域的特征分布的 MMD 距离很大. 经过预训练后, 其 MMD 距离值会减

少,最后经过半监督训练可以再次降低 MMD 距离值. 实验结果说明了,FSLSS 模型可以更有效地拉近源域和

目标域之间的特征分布距离,使模型迁移效果更好,从而提高模型的泛化能力.

表 4 源域和目标域的特征分布的 MMD 距离值

模型	5 way 1 shot(单位 10^{-2})				5 way 5 shot(单位 10^{-2})			
	第 1 次	第 2 次	第 3 次	第 4 次	第 1 次	第 2 次	第 3 次	第 4 次
未训练	9.30	8.48	10.13	13.16	92.24	90.15	93.56	85.04
预训练	7.23	6.22	6.90	7.30	33.91	33.06	33.35	33.49
半监督训练	6.20	5.15	6.24	6.2	31.59	32.24	32.10	32.59

5 总结

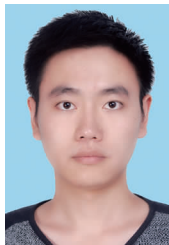
少样本学习已成为机器学习研究领域的难点与热点.它只需要很少的标记样本,就能建立识别率较高的机器学习模型.本文针对现有少样本学习算法在源域和目标域样本分布差异较大的情况下,训练出来的模型泛化能力弱的问题,基于半监督学习、深度网络,提出了一种 FSLSS 模型的少样本学习方法.通过经典少样本数据集上的对比实验,证明了 FSLSS 模型具有较好的泛化能力;同时,利用 MMD 距离评估 FSLSS 模型的迁移效果,表明了 FSLSS 模型可拉近源域数据和目标域数据的特征分布距离,使模型迁移效果更好.

参考文献

- [1] Yann L, Yoshua B, Geoffrey Hinton. Deep learning [J]. Nature, 2015, 521 (7553): 436 - 444.
- [2] Lin X, Rivenson Y, Yardimci N T, et al. All-optical machine learning using diffractive deep neural networks [J]. Science, 2018, 361 (6406): 1004 - 1008.
- [3] Fazeli N, Oller M, Wu J, et al. See, feel, act: Hierarchical learning for complex manipulation skills with multisensory fusion [J]. Science Robotics, 2019, 4 (26): eaav3123.
- [4] 李小薪, 梁荣华. 有遮挡人脸识别综述: 从子空间回归到深度学习 [J]. 计算机学报, 2018, 41 (01): 177 - 207.
Li Xiao-shui, Liang Ronghua. Overview of occluded face recognition: from subspace regression to deep learning [J]. Chinese Journal of Computers, 2018, 41 (01): 177 - 207. (in Chinese)
- [5] 赵嘉, 等. 深度学习萤火虫算法 [J]. 电子学报, 2018, 46 (11): 2633 - 2641.
Zhao Jia, et al. Deep learning firefly algorithm [J]. Acta Electronica Sinica, 2018, 46 (11): 2633 - 2641. (in Chinese)
- [6] Pan S J, Yang Q. A survey on transfer learning [J]. IEEE Transactions on Knowledge and Data Engineering, 2010, 22 (10): 1345 - 1359.
- [7] Li FF, Robert F, Pietro P. A bayesian approach to unsupervised one-shot learning of object categories [A]. Ninth IEEE International Conference on Computer Vision [C]. Sydney, Australia; IEEE, 2003. 1134 - 1141.
- [8] Jake S, Kevin S, Ricard S Z. Prototypical networks for few-shot learning [A]. Neural Information Processing Systems [C]. Long Beach, California, 2017.
- [9] Maas A, Kemp C. One-shot learning with bayesian networks [A]. 31st Annual Conference of the Cognitive Science Society [C]. Amsterdam, The Netherlands, 2009.
- [10] 舒醒, 等. 基于边际 Fisher 准则和迁移学习的小样本集分类器设计算法 [J]. 自动化学报, 2016, 42 (9): 1313 - 1321.
Shu Xing, et al. Smallsample set classifier design algorithm based on marginal fisher criterion and migration learning [J]. Acta Automatica Sinica, 2016, 42 (9): 1313 - 1321. (in Chinese)
- [11] Gregory K, Richard Z, Ruslan S. Siamese neural networks for one-shot image recognition [A]. International Conference on Machine Learning Deep Learning workshop [C]. Lille, France, 2015.
- [12] Oriol V, Charles B, Timothy L, et al. Matching networks for one shot learning [A]. Advances in Neural Information Processing Systems [C]. Barcelona, Spain, 2016. 3630 - 3638.
- [13] Flood S, YANG Y X, Zhang L, et al. Learning to compare: relation network for few-shot learning [Z/DB]. arXiv preprint arXiv:1711.06025v2, 2018.
- [14] Adam S Sergey B, Matthew B M. One-shot learning with memory-augmented neural networks [A]. International Conference on Machine Learning [C]. New York, USA, 2016.
- [15] Sachin R, Hugo L. Optimization as a model for few-shot learning [A]. International Conference on Learning Representations [C]. Toulon, France, 2017.
- [16] Chelsea F, Pieter A, Sergey L. Model-agnostic meta-learning for fast adaptation of deep networks [A]. International Conference on Machine Learning [C]. Sydney, Australia, 2017.
- [17] Zhou Z H, Li M. Semi-supervised learning by disagreement [J]. Knowledge and Information Systems, 2010, 24 (3): 415 - 439.
- [18] 刘建伟, 刘媛, 罗雄麟. 半监督学习方法 [J]. 计算机学报, 2015, 38 (08): 1592 - 1617.

- Liu Jianwei, et al. Semi-supervised learning method [J]. Chinese Journal of Computers, 2015, 38 (08): 1592 – 1617. (in Chinese)
- [19] Ben D, Lu T, Pai D. Does unlabeled data provably help? Worst-case analysis of the sample complexity of semi-supervised learning [A]. Proceedings of the 21st Annual Conference on Learning Theory [C]. Helsinki, Finland, 2008. 33 – 44.
- [20] Balcan M F, Blum A. A discriminative model for semi-supervised learning [J]. Journal of the Association for Computing Machinery, 2010, 57(3): 19.
- [21] Hinton G, Salakhutdinov R R. Reducing the dimensionality of data with neural networks [J]. Science, 2006, 313 (5786): 504 – 507.
- [22] Jo Y J, Park S, Jung J H, et al. Holographic deep learning for rapid optical screening of anthrax spores [J]. Science Advances, 2017, 3(8): e1700606.
- [23] Lin X, Rivenson Y, Yardimci N T, et al. All-optical machine learning using diffractive deep neural networks [J]. Science, 2018, 361 (6406): 1004 – 1008.
- [24] Hirschberg J, Manning C D. Advances in natural language processing [J]. Science, 2015, 349 (6245): 261 – 266.
- [25] Alex K, I S, Geoffrey H. Imagenet classification with deep convolutional neural networks [J]. Advances in Neural Information Processing Systems, 2012, 1: 1097 – 1105.
- [26] Lee, Dong H. Pseudo-Label: The simple and efficient semi-supervised learning method for deep neural networks [A]. ICML 2013 Workshop on Challenges in Representation Learning [C]. Paris, France, 2013. 2 – 10.
- [27] Brenden L, Ruslan S, Jason G, et al. One shot learning of simple visual concepts [A]. Cognitive Science Society Annual Conference [C]. Paris, France, 2011.
- [28] Gretton A. A kernel two-sample test [J]. Journal of Machine Learning Research, 2012, 13(1): 723 – 773.

作者简介



余 游 男, 1993 年出生于四川遂宁, 现为四川师范大学计算机科学学院硕士研究生, 主要研究方向为深度学习与数据挖掘。
E-mail: 454665275@qq.com



冯 林 (通信作者) 男, 1972 年出生于四川巴中, 现为四川师范大学计算机科学学院博士, 教授, 研究生导师, 主要研究方向为数据挖掘。
E-mail: jkxy_fl@sicnu.edu.cn



王格格 女, 1995 年出生于重庆市, 现为四川师范大学计算机科学学院硕士研究生, 主要研究方向为人工智能与数据挖掘。
E-mail: 347673996@qq.com



徐其凤 女, 1993 年出生于四川宜宾, 现为四川师范大学计算机科学学院硕士研究生, 主要研究方向为深度学习。
E-mail: 616876864@qq.com