



(12) 发明专利

(10) 授权公告号 CN 114782779 B

(45) 授权公告日 2023. 06. 02

(21) 申请号 202210487387.2

(22) 申请日 2022.05.06

(65) 同一申请的已公布的文献号
申请公布号 CN 114782779 A

(43) 申请公布日 2022.07.22

(73) 专利权人 兰州理工大学
地址 730050 甘肃省兰州市七里河区兰工
坪路287号

(72) 发明人 李晓旭 王湘阳 刘俊 金志宇
任凯 张文斌 曾俊瑀 李睿凡
陶剑 董洪飞

(74) 专利代理机构 北京挺立专利事务所(普通
合伙) 11265
专利代理师 高福勇

(51) Int.Cl.

G06V 10/774 (2022.01)

G06V 10/764 (2022.01)

G06N 20/00 (2019.01)

(56) 对比文件

CN 101750611 A, 2010.06.23

CN 105469111 A, 2016.04.06

审查员 张杨

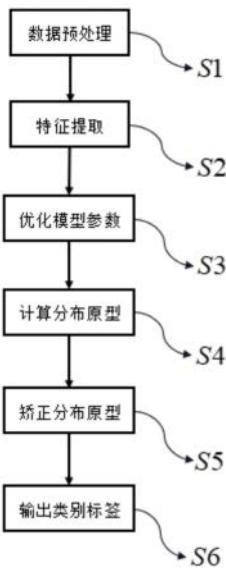
权利要求书4页 说明书9页 附图2页

(54) 发明名称

基于特征分布迁移的小样本图像特征学习
方法及装置

(57) 摘要

本发明公开了一种基于特征分布迁移的小样本图像特征学习方法及装置,在前期利用基类的数据结合梯度下降的方法,对嵌入模块以及分布学习模块的参数进行优化,后期进行分布矫正时,并不需要额外的参数设置;另外,通常假设特征表示中的每个维度都遵循高斯分布,这样高斯分布的均值和方差可以在类似的类别之间传递,减少偏差,以便这些类别的统计数据在足够的样本数下得到更好的估计,再利用分布矫正模型,对样本的分布进行矫正,从而更为精准的对新类样本进行分类。同时可以与任何分类器和特征提取器配对,无需额外的参数,解决了小样本图像分类中存在的原型偏差问题,改善了图像的分类效果,具有很高的实用价值。



1. 一种基于特征分布迁移的小样本图像特征学习方法,其特征在于,包括以下步骤:

S1、对数据进行预处理,其中数据包括训练集和测试集;

S2,利用基类数据预训练嵌入模块 f_0 ,得到特征空间;

S3,将 D_{train} 输入到嵌入模块 f_0 ,得到样本特征图,将其输入到分布学习模块 g_ϕ 中,最小化损失函数,优化分布学习模块 g_ϕ ;

S4,将新类数据分为支持集 \tilde{S}_k 和查询集 \tilde{Q}_k ,将支持集 \tilde{S}_k 经过嵌入模块 f_0 和分布学习模块 g_ϕ 计算每类的分布原型 μ_{s_c} 和 Σ_{s_c} ;

S5,计算基类数据中各类的类别概率,选取最大的前n个类别,将n个类别的分布与当前类别的分布合并,得到矫正后每类的分布原型 $\hat{\mu}_k$ 和 $\hat{\Sigma}_k$;步骤S5具体为:

S51,计算基类样本数据中各个类的类别概率,公式如下:

$$S_d = p_r \left(\mu_{S_c} \mid p(Z_c) \right) \quad (8)$$

式中, $Z_c \sim \mathcal{N}(\mu_c, \sigma_c)$ 表示为基类样本数据中类别C的均值与方差服从高斯分布, μ_{s_c} 表示为支持集 \tilde{S}_k 中的第C类的均值, S_d 表示为将支持集 \tilde{S}_k 第C类的分布作为输入,与基类样本数据中的第C类的分布相比较的距离集;

S52,选取最大的前n个类别,将n个类别的分布与当前类别的分布合并,公式如下:

$$S_N = \left\{ i \mid p_r \left(\mu_{S_c} \mid p(Z_c) \right) \in \text{topn}(S_d) \right\} \quad (9)$$

式中, $\text{topn}(\cdot)$ 表示为一个从输入距离集 S_d 中选择顶部元素的操作符, S_N 用以存储关于特征向量最近的n个最近的基类样本数据;

S53,将合并后的类输入公式(6)和(7),得到矫正后的每个类的分布原型 $\hat{\mu}_k$ 和 $\hat{\Sigma}_k$:

$$\mu_{s_c} = \left\{ \sum_{x_i \in S_c} \sum (x_i)^{-1} \right\}^{-1} \left\{ \sum_{x_i \in S_c} \sum (x_i)^{-1} \mu_{x_i} \right\} \quad (6)$$

$$\Sigma_{s_c} = \left\{ \frac{1}{|S_c|} \sum_{x_i \in S_c} \left[\sum (x_i)^{-1} \right] \right\}^{-1} \quad (7)$$

式(6)中, S_c 表示为支持集 \tilde{S}_k 中的第C类, x_i 表示为支持集 \tilde{S}_k 中的第C类中的样本, μ_{x_i} 表示为样本 x_i 的均值, μ_c 表示为第C类的均值即第C个类的分布,式(6)整体表示为求第C类的加权调和平均数,以表示在模型中不同类的分布原型的位置,用以收紧类内关系和满足识别差距;

式(7)的目的是求类别C的方差,用以在足够的类别信息下消除单个数据的类无关表示,减少整体类别信息的幅度变化;

S6,计算新类查询样本的预测概率。

2. 根据权利要求1所述的基于特征分布迁移的小样本图像特征学习方法,其特征在于,步骤S1的预处理方法为:

S11,将数据 $D = \{(x_i, y_i), y_i \in L\}_{i=1}^N$ 分为 $D_{train} = \{(x_i, y_i), y_i \in L_{train}\}_{i=1}^{N_{train}}$ 和

$D_{test} = \{(x_i^*, y_i^*), y_i^* \in L_{test}\}_{i=1}^{N_{test}}$ 两部分,且这两部分的类别空间互斥,将 D_{train} 用以在训练过程中调整参数, D_{test} 作为新类数据测评模型性能;

S12,对于C-way K-shot分类任务,从 D_{train} 中随机选出C个类别,每个类别中随机选出M个样本,其中K个样本作为支持样本 S_i ,其余M-K个样本作为查询样本 Q_i , S_i 和 Q_i 构成一个任务 T_i ;同样地,对于 D_{test} 有任务 \tilde{T}_k 。

3.根据权利要求1所述的基于特征分布迁移的小样本图像特征学习方法,其特征在于,步骤S2中使用包含四个卷积块的嵌入模块 f_θ 对图像提取特征,其中含有卷积层、池化层和非线性激活函数,每个卷积块使用窗口大小为3*3的卷积核,一个批量归一化,一个RELU非线性层,一个2*2最大池化层,裁剪了最后两个块的最大池化层。

4.根据权利要求1所述的基于特征分布迁移的小样本图像特征学习方法,其特征在于,步骤S3中分布学习模块 g_ϕ 由两个全连接层组成,用以提取图像特征的分布表示,得到类中每个样本的均值与方差。

5.根据权利要求1所述的基于特征分布迁移的小样本图像特征学习方法,其特征在于,步骤S3中最小化损失函数使用的是梯度下降算法,通过不断地调整权重 ω 和偏差 b ,使得损失函数的值变得越来越小。

6.根据权利要求1所述的基于特征分布迁移的小样本图像特征学习方法,其特征在于,步骤S3具体包括:

S31、将基类中的 D_{train} 输入嵌入模块 f_θ 中,依次经过卷积层、池化层和激活函数,得每个类的样本特征图;

S32,计算出每个类样本特征图的均值 μ_c 和方差 σ_c ,与预训练样本特征的空间分布相比较,对嵌入模块 f_θ 的参数加以调整,根据公式(1)和(2)计算每个类的均值 μ_c 和方差 σ_c :

$$\mu_c = \frac{\sum_{i=1}^{n_c} (x_i)}{n_c} \quad (1)$$

$$\sigma_c = \frac{\sum_{i=1}^{n_c} (x_i - \mu_c)^2}{n_c} \quad (2)$$

式中 x_i 表示为基类中C的第i个样本的特征向量, n_c 表示为C类中的样本总数;

S33,将各类样本特征图输入到分布学习模块 g_ϕ ,得到每个样本均值 μ_{x_i} 和方差 σ_{x_i} ,利用高斯分布公式(3)计算每个样本 x_i 类别概率:

$$p(x_i) = N(\mu_{x_i}; \mu_c, \Sigma_c), C=1, \dots, C \quad (3)$$

式中 Σ_c 表示为C类特征的协方差矩阵,其计算公式如公式(4)所示:

$$\Sigma_c = \frac{1}{n_c - 1} \sum_{i=1}^{n_c} (x_i - \mu_c)(x_i - \mu_c)^T \quad (4)$$

S34,利用交叉熵公式最小化损失函数,优化分布学习模块 g_ϕ 参数,公式如(5)所示:

$$L = -\sum_{i=1}^n y \log(p(x_i)) \quad (5)$$

式中 y 表示为一组带有标签的特征向量。

7. 根据权利要求1所述的基于特征分布迁移的小样本图像特征学习方法,其特征在于,步骤S4具体如下:

S41, 将新类数据分为支持集 \tilde{S}_k 和查询集 \tilde{Q}_k , 每个任务 \tilde{T}_k 由支持集 \tilde{S}_k 和查询集 \tilde{Q}_k 组成;

S42, 将支持集 \tilde{S}_k 输入嵌入模块 f_0 , 得到每个类样本特征图的均值 μ_c 和方差 σ_c ;

S43, 将各类样本特征图输入到分布学习模块 g_ϕ , 得到每个样本均值 μ_{x_i} 和方差 σ_{x_i} ;

S44, 根据每个样本的均值 μ_{x_i} 和方差 σ_{x_i} , 利用公式(6)和(7)计算 \tilde{S}_k 中每个类的分布原型 μ_{s_c} 和 Σ_{s_c} ;

$$\mu_{s_c} = \left\{ \sum_{x_i \in S_c} \sum (x_i)^{-1} \right\}^{-1} \left\{ \sum_{x_i \in S_c} \sum (x_i)^{-1} \mu_{x_i} \right\} \quad (6)$$

$$\Sigma_{s_c} = \left\{ \frac{1}{|S_c|} \sum_{x_i \in S_c} \left[\sum (x_i)^{-1} \right] \right\}^{-1} \quad (7)$$

式(6)中, S_c 表示为支持集 \tilde{S}_k 中的第C类, x_i 表示为支持集 \tilde{S}_k 中的第C类中的样本, μ_{x_i} 表示为样本 x_i 的均值, μ_c 表示为第C类的均值即第C个类的分布,式(6)整体表示为求第C类的加权调和平均数,以表示在模型中不同类的分布原型的位置,用以收紧类内关系和满足识别差距;

式(7)的目的是求类别C的方差,用以在足够的类别信息下消除单个数据的类无关表示,减少整体类别信息的幅度变化。

8. 根据权利要求1所述的基于特征分布迁移的小样本图像特征学习方法,其特征在于,步骤S6具体为:

S61, 将新类数据的查询集 \tilde{Q}_k 的样本信息输入嵌入模块 f_0 , 得到每个类样本特征图的均值 μ_c 和方差 σ_c ;

S62, 将各类样本特征图输入到分布学习模块 g_ϕ , 得到每个样本均值 μ_{x_i} 和方差 σ_{x_i} ;

S63, 将每个样本的均值 μ_{x_i} 和方差 σ_{x_i} 输入公式(3), 计算出新类查询样本的预测概率, 将其输入到度量模块, 输出对应的类别标签:

$$p(x_i) = N(\mu_{x_i}; \mu_c, \Sigma_c), C=1, \dots, C \quad (3)$$

式中 Σ_c 表示为C类特征的协方差矩阵,其计算公式如公式(4)所示:

$$\Sigma_c = \frac{1}{n_c - 1} \sum_{i=1}^{n_c} (x_i - \mu_c)(x_i - \mu_c)^T \quad (4)。$$

9. 一种面向小样本图像的任务自适应度量学习装置,其特征在于,用以实现权利要求1-8任一项所述的面向小样本图像的任务自适应度量学习方法,包括以下模块:

嵌入模块,用于对图像样本进行特征提取处理,构造特征空间,其中,所述图像样本包

括基类样本、新类支持样本和查询样本；

分布学习模块，用于提取图像特征的分布表示，得到类中每个样本的均值与方差；

分布矫正模块，目的在于用基类样本的分布对新类样本进行分布矫正，构建图像分布矫正模型；

度量模块，用于利用优化后的基类样本的分布对新类查询集样本进行分类，获取类别标签。

基于特征分布迁移的小样本图像特征学习方法及装置

技术领域

[0001] 本发明涉及图像分类技术领域,尤其涉及一种基于特征分布迁移的小样本图像特征学习方法及装置。

背景技术

[0002] 近年来,随着计算机技术的发展,人们浏览的信息日益丰富,每天都有大量图片被上传到网络,由于数量巨大,人工已经无法对此进行分类。在很多大样本图像分类任务上,机器的识别性能已经超越人类。然而,当样本量比较少时,机器的识别水平仍与人类存在较大差距。因此,研究高效可靠的图片分类算法有很迫切的社会需求。

[0003] 小样本分类(Few-shot Classification)属于小样本学习(Few-shot Learning)范畴,往往包含类别空间不相交的两类数据,即基类数据和新类数据。小样本分类旨在利用基类数据学习的知识和新类数据的少量标记样本(支持样本)来学习分类规则,准确预测新类任务中未标记样本(查询样本)的类别。

[0004] 针对现有的技术缺点而言,首先,对标记样本极少的小样本分类任务来说,现有的深度学习技术并不适用。因而,如何基于基类数据和标记样本极少的新类数据,来学习高辨识度的特征表示,是一个值得探索的问题。其次,对分布原型的偏差而言,由于标记样本极少,常常使得学习的原型偏差(bias)较大。因此,如何通过减少原型偏差来提高小样本图像分类性能,也是一项具有挑战的任务。最后,在基类数据特征的判别性以及基类数据特征在新类数据上的可迁移性存在着误差,易导致分类的不准确性。

发明内容

[0005] 本发明针对上述小样本图像分类中的原型偏差问题,提出一种基于特征分布迁移的小样本图像特征学习方法及装置,主要是通常假设特征表示中的每个维度都遵循高斯分布,这样高斯分布的均值和方差可以在类似的类别之间传递,结合比较样本间或者样本与分布原型间的距离来判断类别,这些类别的统计数据在足够的样本数下得到更好的估计,改善了图像的分类效果。

[0006] 为了实现上述目的,本发明提供如下技术方案:

[0007] 一方面,本发明提供了一种基于特征分布迁移的小样本图像特征学习方法,包括以下步骤:

[0008] S1、对数据进行预处理,其中数据包括训练集和测试集;

[0009] S2,利用基类数据预训练嵌入模块 f_0 ,得到良好的特征空间;

[0010] S3,将 D_{train} 输入到嵌入模块 f_0 ,得到样本特征图,将其输入到分布学习模块 g_ϕ 中,最小化损失函数,优化分布学习模块 g_ϕ ;

[0011] S4,将新类数据分为支持集 \tilde{S}_k 和查询集 \tilde{Q}_k ,将支持集 \tilde{S}_k 经过嵌入模块 f_0 和分布学习模块 g_ϕ 计算每类的分布原型 μ_{s_k} 和 Σ_{s_k} ;

[0012] S5,计算基类数据中各类的类别概率,选取最大的前n个类别,将n个类别的分布与当前类别的分布合并,得到矫正后每类的分布原型 $\hat{\mu}_k$ 和 $\hat{\Sigma}_k$;

[0013] S6,计算新类查询样本的预测概率。

[0014] 进一步地,步骤S1的预处理方法为:

[0015] S 1 1 , 将 数 据 $D = \{(x_i, y_i), y_i \in L\}_{i=1}^N$ 分 为 $D_{train} = \{(x_i, y_i), y_i \in L_{train}\}_{i=1}^{N_{train}}$ 和 $D_{test} = \{(x_i^*, y_i^*), y_i^* \in L_{test}\}_{i=1}^{N_{test}}$ 两部分,且这两部分的类别空间互斥,将 D_{train} 用以在训练过程中调整参数, D_{test} 作为新类数据测评模型性能;

[0016] S12,对于C-way K-shot分类任务,从 D_{train} 中随机选出C个类别,每个类别中随机选出M个样本,其中K个样本作为支持样本 S_i ,其余M-K个样本作为查询样本 Q_i , S_i 和 Q_i 构成一个任务 T_i ;同样地,对于 D_{test} 有任务 \tilde{T}_k 。

[0017] 进一步地,步骤S2中使用包含四个卷积块的嵌入模块 f_θ 对图像提取特征,其中含有卷积层、池化层和非线性激活函数,每个卷积块使用窗口大小为3*3的卷积核,一个批量归一化,一个RELU非线性层,一个2*2最大池化层,裁剪了最后两个块的最大池化层。

[0018] 进一步地,步骤S3中分布学习模块 g_ϕ 由两个全连接层组成,用以提取图像特征的分布表示,得到类中每个样本的均值与方差。

[0019] 进一步地,步骤S3中最小化损失函数使用的是梯度下降算法,不断地调整权重 ω 和偏差 b ,使得损失函数的值变得越来越小。

[0020] 进一步地,步骤S3具体包括:

[0021] S31、将基类中的 D_{train} 输入嵌入模块 f_θ 中,依次经过卷积层、池化层和激活函数,得每个类的样本特征图;

[0022] S32,计算出每个类样本特征图的均值 μ_c 和方差 σ_c ,与预训练样本特征的空间分布相比较,对嵌入模块 f_θ 的参数加以调整,根据公式(1)和(2)计算每个类的均值 μ_c 和方差 σ_c :

$$[0023] \quad \mu_c = \frac{\sum_{i=1}^{n_c} (x_i)}{n_c} \quad (1)$$

$$[0024] \quad \sigma_c = \frac{\sum_{i=1}^{n_c} (x_i - \mu_c)^2}{n_c} \quad (2)$$

[0025] 式中 x_i 表示为基类中C的第i个样本的特征向量, n_c 表示为C类中的样本总数;

[0026] S33,将各类样本特征图输入到分布学习模块 g_ϕ ,得到每个样本均值 μ_{x_i} 和方差 σ_{x_i} ,利用高斯分布公式(3)计算每个样本 x_i 类别概率:

$$[0027] \quad p(x_i) = N(\mu_{x_i}; \mu_c, \Sigma_c), C=1, \dots, C \quad (3)$$

[0028] 式中 Σ_c 表示为C类特征的协方差矩阵,其计算公式如公式(4)所示:

$$[0029] \quad \Sigma_c = \frac{1}{n_c - 1} \sum_{i=1}^{n_c} (x_i - \mu_c)(x_i - \mu_c)^T \quad (4)$$

[0030] S34,利用交叉熵公式最小化损失函数,优化分布学习模块 g_ϕ 参数,公式如(5)所示:

[0031]
$$L = - \sum_{i=1}^n y \log(p(x_i)) \quad (5)$$

[0032] 式中 y 表示为一组带有标签的特征向量。

[0033] 进一步地,步骤4具体如下:

[0034] S41,每个任务 \tilde{T}_k 由支持集 \tilde{S}_k 和查询集 \tilde{Q}_k 组成;

[0035] S42,将支持集 \tilde{S}_k 输入嵌入模块 f_θ ,得到每个类样本特征图的均值 μ_c 和方差 σ_c ;

[0036] S43,将各类样本特征图输入到分布学习模块 g_ϕ ,得到每个样本均值 μ_{x_i} 和方差 σ_{x_i} ;

[0037] S44,根据每个样本的均值 μ_{x_i} 和方差 σ_{x_i} ,利用公式(6)和(7)计算 \tilde{S}_k 中每个类的分布原型 μ_{s_c} 和 Σ_{s_c} ;

[0038]
$$\mu_{s_c} = \left\{ \sum_{x_i \in S_c} \sum (x_i)^{-1} \right\}^{-1} \left\{ \sum_{x_i \in S_c} \sum (x_i)^{-1} \mu_{x_i} \right\} \quad (6)$$

[0039]
$$\Sigma_{s_c} = \left\{ \frac{1}{|S_c|} \sum_{x_i \in S_c} \left[\sum (x_i)^{-1} \right] \right\}^{-1} \quad (7)$$

[0040] 式(6)中, S_c 表示为支持集 \tilde{S}_k 中的第C类, x_i 表示为支持集 \tilde{S}_k 中的第C类中的样本, μ_{x_i} 表示为样本 x_i 的均值, μ_c 表示为第C类的均值即第C个类的分布,式(6)整体表示为求第C类的加权调和平均数,以表示在模型中不同类的分布原型的位置,用以收紧类内关系和满足识别差距;

[0041] 式(7)的目的是求类别C的方差,用以在足够的类别信息下消除单个数据的类无关表示,减少整体类别信息的幅度变化。

[0042] 进一步地,步骤S5具体为:

[0043] S51,计算基类样本数据中各个类的类别概率,公式如下:

[0044]
$$S_d = p_r(\mu_{S_c} | p(Z_c)) \quad (8)$$

[0045] 式中, $Z_c \sim \mathcal{N}(\mu_c, \sigma_c)$ 表示为基类样本数据中类别C的均值与方差服从高斯分布, μ_{S_c} 表示为支持集 \tilde{S}_k 中的第C类的均值, S_d 表示为将支持集 \tilde{S}_k 第C类的分布作为输入,与基类样本数据中的第C类的分布相比较的距离集;

[0046] S52,选取最大的前n个类别,将n个类别的分布与当前类别的分布合并,公式如下:

[0047]
$$S_N = \left\{ i | p_r(\mu_{S_c} | p(Z_c)) \in \text{topn}(S_d) \right\} \quad (9)$$

[0048] 式中, $\text{topn}(\cdot)$ 表示为一个从输入距离集 S_d 中选择顶部元素的操作符, S_N 用以存储关于特征向量最近的n个最近的基类样本数据;

[0049] S53,将合并后的类输入公式(6)和(7),得到矫正后的每个类的分布原型 $\hat{\mu}_k$ 和 $\hat{\Sigma}_k$:

$$[0050] \quad \mu_{s_c} = \left\{ \sum_{x_i \in S_c} \sum (x_i)^{-1} \right\}^{-1} \left\{ \sum_{x_i \in S_c} \sum (x_i)^{-1} \mu_{x_i} \right\} \quad (6)$$

$$[0051] \quad \Sigma_{s_c} = \left\{ \frac{1}{|S_c|} \sum_{x_i \in S_c} \left[\sum (x_i)^{-1} \right] \right\}^{-1} \quad (7)$$

[0052] 式(6)中, S_c 表示为支持集 \tilde{S}_k 中的第C类, x_i 表示为支持集 \tilde{S}_k 中的第C类中的样本, μ_{x_i} 表示为样本 x_i 的均值, μ_c 表示为第C类的均值即第C个类的分布,式(6)整体表示为求第C类的加权调和平均数,以表示在模型中不同类的分布原型的位置,用以收紧类内关系和满足识别差距;

[0053] 式(7)的目的是求类别C的方差,用以在足够的类别信息下消除单个数据的类无关表示,减少整体类别信息的幅度变化。

[0054] 进一步地,步骤S6具体为:

[0055] S61,将新类数据的查询集 \tilde{Q}_k 的样本信息输入嵌入模块 f_0 ,得到每个类样本特征图的均值 μ_c 和方差 σ_c ;

[0056] S62,将各类样本特征图输入到分布学习模块 g_ϕ ,得到每个样本均值 μ_{x_i} 和方差 σ_{x_i} ;

[0057] S63,将每个样本的均值 μ_{x_i} 和方差 σ_{x_i} 输入公式(3),计算出新类查询样本的预测概率,将其输入到度量模块,输出对应的类别标签:

$$[0058] \quad p(x_i) = N(\mu_{x_i}; \mu_c, \Sigma_c), C=1, \dots, C \quad (3)$$

[0059] 式中 Σ_c 表示为C类特征的协方差矩阵,其计算公式如公式(4)所示:

$$[0060] \quad \Sigma_c = \frac{1}{n_c - 1} \sum_{i=1}^{n_c} (x_i - \mu_c)(x_i - \mu_c)^T \quad (4)。$$

[0061] 另一方面,本发明还提供了一种面向小样本图像的任务自适应度量学习装置,用以实现上述的任一项方法,包括以下模块:

[0062] 嵌入模块,用于对图像样本进行特征提取处理,构造特征空间,其中,所述图像样本包括基类样本、新类支持样本和查询样本;

[0063] 分布学习模块,用于提取图像特征的分布表示,得到类中每个样本的均值与方差;

[0064] 分布矫正模块,目的在于用基类样本的分布对新类样本进行分布矫正,构建图像分布矫正模型;

[0065] 度量模块,用于利用优化后的基类样本的分布对新类查询集样本进行分类,获取类别标签。

[0066] 与现有技术相比,本发明的有益效果为:

[0067] 本发明建立了一种基于特征分布迁移的小样本图像特征学习方法及装置,可以与任何分类器和特征提取器配对,无需额外的参数,解决了小样本图像分类中存在的原型偏差问题,改善了图像的分类效果,具有很高的实用价值。

[0068] 本发明的装置在前期利用基类的数据结合梯度下降的方法,对嵌入模块以及分布学习模块的参数进行优化,后期进行分布矫正时,并不需要额外的参数设置;另外,通常假

设特征表示中的每个维度都遵循高斯分布,这样高斯分布的均值和方差可以在类似的类别之间传递,减少偏差,以便这些类别的统计数据在足够的样本数下得到更好的估计,再利用分布矫正模型,对样本的分布进行矫正,从而更为精准的对新类样本进行分类。

附图说明

[0069] 为了更清楚地说明本申请实施例或现有技术中的技术方案,下面将对实施例中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本发明中记载的一些实施例,对于本领域普通技术人员来讲,还可以根据这些附图获得其他的附图。

[0070] 图1为本发明实施例提供的基于特征分布迁移的小样本图像特征学习方法流程图。

[0071] 图2为本发明实施例提供的基于特征分布迁移的小样本图像特征学习模型的迁移学习和分布迁移的特征学习网络结构图。

[0072] 图3为本发明实施例提供的分布矫正模块的流程图。

[0073] 图4为本发明实施例提供的基于特征分布迁移的小样本图像特征学习装置功能模块示意图。

具体实施方式

[0074] 下面结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例是本发明一部分实施例,而不是全部的实施例。本发明中的实施例,本领域技术人员在没有做出创造性劳动前提下所获得的所有其他实施例,都属于本发明保护的范围。

[0075] 根据本文公开的一个方面,提供了一种基于特征分布迁移的小样本图像特征学习方法,如图1所示,包括以下阶段步骤:

[0076] S1、对数据进行预处理,其中数据包括训练集和测试集;

[0077] 具体地,步骤S1的预处理方法包括:

[0078] S11,将数据 $D = \{(x_i, y_i), y_i \in L\}_{i=1}^N$ 分为 $D_{train} = \{(x_i, y_i), y_i \in$

[0079] $L_{train}\}_{i=1}^{N_{train}}$ 和 $D_{test} = \{(x_i^*, y_i^*), y_i^* \in L_{test}\}_{i=1}^{N_{test}}$ 两部分,且这两部分的类别空间互斥,将 D_{train} 用以在训练过程中调整参数, D_{test} 作为新类数据测评模型性能;

[0080] S12,对于C-way K-shot分类任务,从 D_{train} 中随机选出C个类别,每个类别中随机选出M个样本,其中K个样本作为支持样本 S_i ,其余M-K个样本作为查询样本 Q_i , S_i 和 Q_i 构成一个任务 T_i ;同样地,对于 D_{test} 有任务 \tilde{T}_k 。

[0081] S2,利用基类数据预训练嵌入模块 f_0 ,得到良好的特征空间;

[0082] 进一步地,步骤S2中使用包含四个卷积块的嵌入模块 f_0 对图像提取特征,其中含有卷积层、池化层和非线性激活函数,每个卷积块使用窗口大小为3*3的卷积核,一个批量归一化,一个RELU非线性层,一个2*2最大池化层,裁剪了最后两个块的最大池化层。例如,对于84*84*3RGB图像,每个块使用一个带有64个滤波器的3x3的卷积核。

[0083] S3,将 D_{train} 输入到嵌入模块 f_0 ,得到样本特征图,将其输入到分布学习模块 g_ϕ 中,最小化损失函数,优化分布学习模块 g_ϕ ;

[0084] 其中,分布学习模块 g_ϕ 由两个全连接层组成,用以提取图像特征的分布表示,得到类中每个样本的均值与方差。

[0085] 最小化损失函数使用的是梯度下降算法,不断地调整权重 ω 和偏差 b ,使得损失函数的值变得越来越小。也可以替换为随机梯度下降或者批量梯度下降等。

[0086] 具体地,步骤S3包括:

[0087] S31、将基类中的 D_{train} 输入嵌入模块 f_θ 中,依次经过卷积层、池化层和激活函数,得每个类的样本特征图;

[0088] S32,计算出每个类样本特征图的均值 μ_c 和方差 σ_c ,与预训练样本特征的空间分布相比较,对嵌入模块 f_θ 的参数加以调整,根据公式(1)和(2)计算每个类的均值 μ_c 和方差 σ_c :

$$[0089] \quad \mu_c = \frac{\sum_{i=1}^{n_c} (x_i)}{n_c} \quad (1)$$

$$[0090] \quad \sigma_c = \frac{\sum_{i=1}^{n_c} (x_i - \mu_c)^2}{n_c} \quad (2)$$

[0091] 式中 x_i 表示为基类中C的第i个样本的特征向量, n_c 表示为C类中的样本总数;

[0092] S33,将各类样本特征图输入到分布学习模块 g_ϕ ,得到每个样本均值 μ_{x_i} 和方差 σ_{x_i} ,利用高斯分布公式(3)计算每个样本 x_i 类别概率:

$$[0093] \quad p(x_i) = N(\mu_{x_i}; \mu_c, \Sigma_c), C=1, \dots, C \quad (3)$$

[0094] 式中 Σ_c 表示为C类特征的协方差矩阵,其计算公式如公式(4)所示:

$$[0095] \quad \Sigma_c = \frac{1}{n_c - 1} \sum_{i=1}^{n_c} (x_i - \mu_c)(x_i - \mu_c)^T \quad (4)$$

[0096] S34,利用交叉熵公式最小化损失函数,优化分布学习模块 g_ϕ 参数,公式如(5)所示:

$$[0097] \quad L = - \sum_{i=1}^n y \log(p(x_i)) \quad (5)$$

[0098] 式中 y 表示为一组带有标签的特征向量。

[0099] S4,将支持集 \tilde{S}_k 经过嵌入模块 f_θ 和分布学习模块 g_ϕ 计算每类的分布原型 μ_{s_c} 和 Σ_{s_c} ;

[0100] 具体地,步骤4包括:

[0101] S41,将新类数据分为支持集 \tilde{S}_k 和查询集 \tilde{Q}_k ,每个任务 \tilde{T}_k 由支持集 \tilde{S}_k 和查询集 \tilde{Q}_k 组成;

[0102] S42,将支持集 \tilde{S}_k 输入嵌入模块 f_θ ,得到每个类样本特征图的均值 μ_c 和方差 σ_c ;

[0103] S43,将各类样本特征图输入到分布学习模块 g_ϕ ,得到每个样本均值 μ_{x_i} 和方差 σ_{x_i} ;

[0104] S44,根据每个样本的均值 μ_{x_i} 和方差 σ_{x_i} ,利用公式(6)和(7)计算 \tilde{S}_k 中每个类的分布原型 μ_{s_c} 和 Σ_{s_c} ;

$$[0105] \quad \mu_{s_c} = \left\{ \sum_{x_i \in S_c} \sum (x_i)^{-1} \right\}^{-1} \left\{ \sum_{x_i \in S_c} \sum (x_i)^{-1} \mu_{x_i} \right\} \quad (6)$$

$$[0106] \quad \Sigma_{s_c} = \left\{ \frac{1}{|S_c|} \sum_{x_i \in S_c} \left[\sum (x_i)^{-1} \right] \right\}^{-1} \quad (7)$$

[0107] 式(6)中, S_c 表示为支持集 \tilde{S}_k 中的第C类, x_i 表示为支持集 \tilde{S}_k 中的第C类中的样本, μ_{x_i} 表示为样本 x_i 的均值, μ_c 表示为第C类的均值即第C个类的分布, 式(6)整体表示为求第C类的加权调和平均数, 以表示在模型中不同类的分布原型的位置, 用以收紧类内关系和满足识别差距;

[0108] 式(7)的目的是求类别C的方差, 用以在足够的类别信息下消除单个数据的类无关表示, 减少整体类别信息的幅度变化。

[0109] S5, 计算基类数据中各类的类别概率, 选取最大的前n个类别, 将n个类别的分布与当前类别的分布合并, 得到矫正后每类的分布原型 $\hat{\mu}_k$ 和 $\hat{\Sigma}_k$;

[0110] 具体地, 步骤S5包括:

[0111] S51, 计算基类样本数据中各个类的类别概率, 公式如下:

$$[0112] \quad S_d = p_r \left(\mu_{S_c} \mid p(Z_c) \right) \quad (8)$$

[0113] 式中, $Z_c \sim \mathcal{N}(\mu_c, \sigma_c)$ 表示为基类样本数据中类别C的均值与方差服从高斯分布, μ_{S_c} 表示为支持集 \tilde{S}_k 中的第C类的均值, S_d 表示为将支持集 \tilde{S}_k 第C类的分布作为输入, 与基类样本数据中的第C类的分布相比较的距离集;

[0114] S52, 选取最大的前n个类别, 将n个类别的分布与当前类别的分布合并, 公式如下:

$$[0115] \quad S_N = \left\{ i \mid p_r \left(\mu_{S_c} \mid p(Z_c) \right) \in \text{topn}(S_d) \right\} \quad (9)$$

[0116] 式中, $\text{topn}(\cdot)$ 表示为一个从输入距离集 S_d 中选择顶部元素的操作符, S_N 用以存储关于特征向量最近的n个最近的基类样本数据;

[0117] S53, 将合并后的类输入公式(6)和(7), 得到矫正后的每个类的分布原型 $\hat{\mu}_k$ 和 $\hat{\Sigma}_k$:

$$[0118] \quad \mu_{s_c} = \left\{ \sum_{x_i \in S_c} \sum (x_i)^{-1} \right\}^{-1} \left\{ \sum_{x_i \in S_c} \sum (x_i)^{-1} \mu_{x_i} \right\} \quad (6)$$

$$[0119] \quad \Sigma_{s_c} = \left\{ \frac{1}{|S_c|} \sum_{x_i \in S_c} \left[\sum (x_i)^{-1} \right] \right\}^{-1} \quad (7)$$

[0120] 式(6)中, S_c 表示为支持集 \tilde{S}_k 中的第C类, x_i 表示为支持集 \tilde{S}_k 中的第C类中的样本, μ_{x_i} 表示为样本 x_i 的均值, μ_c 表示为第C类的均值即第C个类的分布, 式(6)整体表示为求第C类的加权调和平均数, 以表示在模型中不同类的分布原型的位置, 用以收紧类内关系和满足识别差距;

[0121] 式(7)的目的是求类别C的方差,用以在足够的类别信息下消除单个数据的类无关表示,减少整体类别信息的幅度变化。

[0122] S6,计算新类查询样本的预测概率,输出类别标签。

[0123] 具体地,步骤S6包括:

[0124] S61,将新类数据的查询集 \tilde{Q}_k 的样本信息输入嵌入模块 f_θ ,得到每个类样本特征图的均值 μ_c 和方差 σ_c ;

[0125] S62,将各类样本特征图输入到分布学习模块 g_ϕ ,得到每个样本均值 μ_{x_i} 和方差 σ_{x_i} ;

[0126] S63,将每个样本的均值 μ_{x_i} 和方差 σ_{x_i} 输入公式(3),计算出新类查询样本的预测概率,将其输入到度量模块,输出对应的类别标签:

$$[0127] \quad p(x_i) = N(\mu_{x_i}; \mu_c, \Sigma_c), C=1, \dots, C \quad (3)$$

[0128] 式中 Σ_c 表示为C类特征的协方差矩阵,其计算公式如公式(4)所示:

$$[0129] \quad \Sigma_c = \frac{1}{n_c - 1} \sum_{i=1}^{n_c} (x_i - \mu_c)(x_i - \mu_c)^T \quad (4)。$$

[0130] 另一方面,本发明还提供了一种面向小样本图像的任务自适应度量学习装置,用以实现上述的任一项方法,包括以下模块:

[0131] 嵌入模块,用于对图像样本进行特征提取处理,构造特征空间,其中,所述图像样本包括基类样本、新类支持样本和查询样本;

[0132] 分布学习模块,用于提取图像特征的分布表示,得到类中每个样本的均值与方差;

[0133] 分布矫正模块,目的在于用基类样本的分布对新类样本进行分布矫正,构建图像分布矫正模型;

[0134] 度量模块,用于利用优化后的基类样本的分布对新类查询集样本进行分类,获取类别标签。

[0135] 本发明基于特征分布迁移的小样本图像特征学习方法,建立在现成的预训练特征抽取器和分类模型之上,可以与任何分类器和特征提取器配对,无需额外的参数;基于特征分布迁移的小样本图像特征学习方法装置中,采用分布学习模块,通常假设特征表示中的每个维度都遵循高斯分布,这样高斯分布的均值和方差可以在类似的类别之间传递,通过计算每个样本的均值和方差,计算出对应类别的加权调和平均数,用以表示模型中不同类别的分布原型的位置,比较与每个类别代表的距离,学会了对样本进行分类。目的在于通过融合与查询样本相连样本特征及它们之间距离的相似性,更新查询样本并分类。

[0136] 以上结合附图对所提出的基于特征分布迁移的小样本图像特征学习方法及模型的具体实施方式进行了阐述。通过以上实施方式的描述,所属领域的技术人员可以清楚的了解该方法以及装置的实施。

[0137] 需要说明的是,在附图和说明书正文中,未描述的实现方式,均为所属技术领域中普通技术人员所知的形式,未进行详细说明。此外,上述对各元件和方法的定义并不仅限于实例中提到的各种具体结构、形状或方式,本领域普通技术人员可对其进行简单地更改或替换。

[0138] 此外,除非特别描述或必须依序发生地步骤,上述步骤地顺序并无限制于以上所列,且可根据所需设计而变化或重新安排。并且上述实例可基于设计及可靠度地考虑,彼此混合搭配使用或与其他实例混合搭配使用,即不同实施中的技术特征可以自由组合形成更多地实施例子。在此提供的算法和显示不与任何特定计算机、虚拟系统或者其他设备固有相关。各种通用系统也可以与基于在此地启示一起使用。根据上面的描述,构造这类系统所要求的结构是显而易见的。此外,本文公开的也不针对任何特定的编程语言。但是应当了解,可以利用各种编程语言实现在此描述的本文公开的内容,并且上面对特定语言所做的描述是为了披露本文公开的最佳实施方式。

[0139] 类似的,应当理解,为了使本文尽量精简并且帮助理解各个公开方面中的一个或多个,在上面对本文公开的示例性实施例的描述中,本文公开的各个特征有时被一起分组到单个实施例、图、或者对其的描述中。然而,并不应将该公开的方法解释成反映如下示意图:即要求所保护的本文公开的要求比在每个权力要求中所明确记载的特征具有更多的特征。更确切地说,如下面的权力要求书所反映的那样,公开方面在于少于前面公开的单个实施例的所有特征。因此,遵循具体实施方式的权利要求书由此明确地并入该具体实施方式,其中每个权利要求本身都作为本公开的单独实施例子。

[0140] 以上所述实施例,仅为本申请的具体实施方式,用以说明本申请的技术方案,而非对其限制,本申请的保护范围并不局限于此,尽管参照前述实施例对本申请进行了详细的说明,本领域的普通技术人员应当理解:任何熟悉本技术领域的技术人员在本申请揭露的技术范围内,其依然可以对前述实施例所记载的技术方案进行修改或可轻易想到变化,或者对其中部分技术特征进行等同替换;而这些修改、变化或者替换,并不使相应技术方案的本质脱离本申请实施例技术方案的精神和范围。都应涵盖在本申请的保护范围之内。因此,本申请的保护范围应所述以权利要求的保护范围为准。

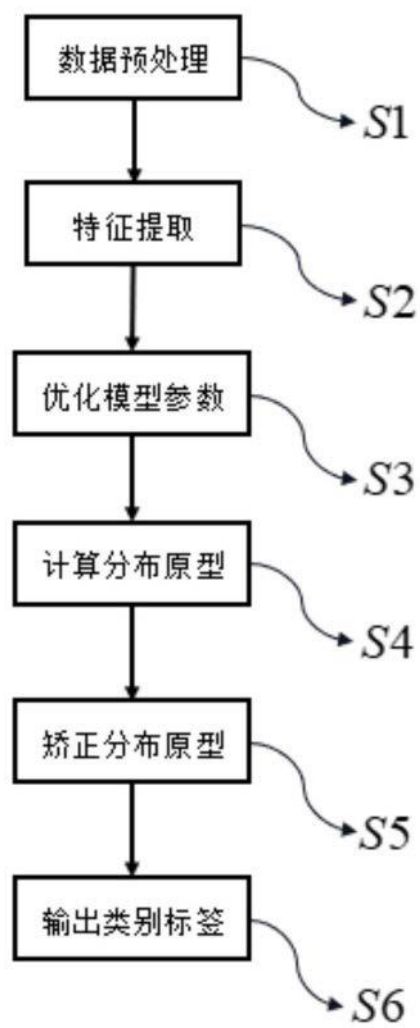


图1

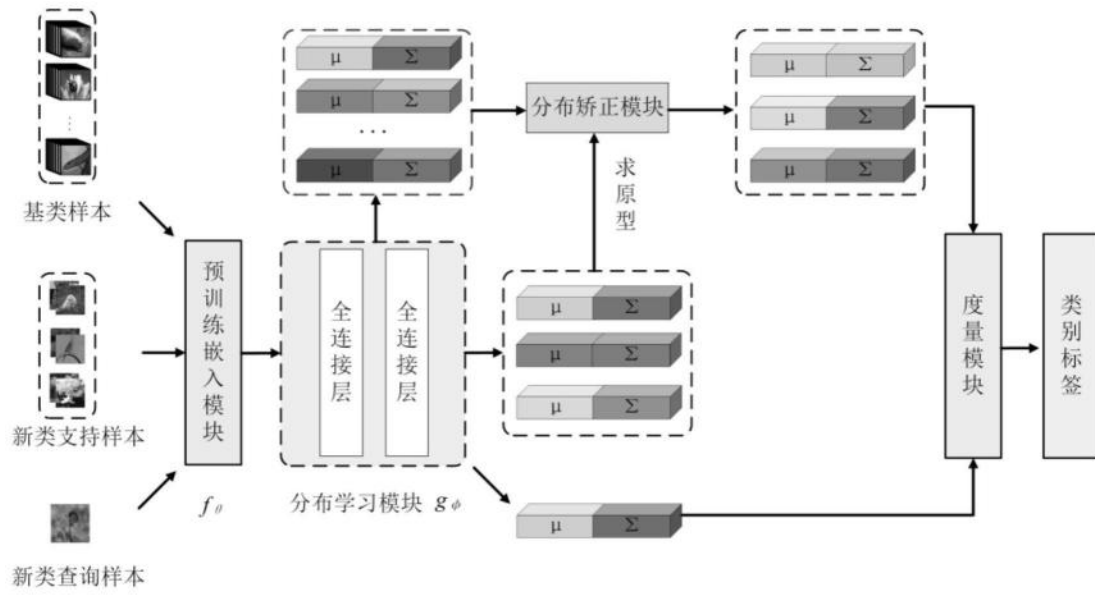


图2

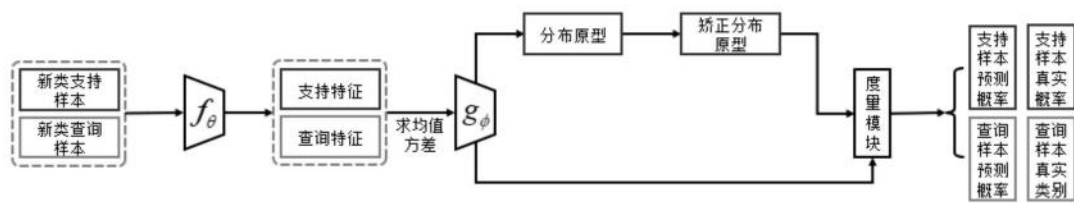


图3

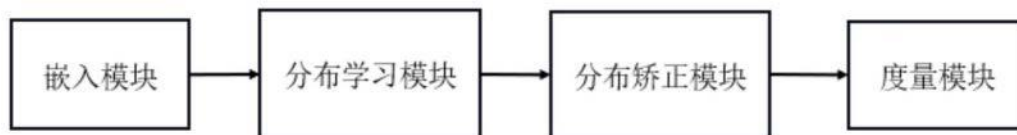


图4