



(12) 发明专利

(10) 授权公告号 CN 114943859 B

(45) 授权公告日 2023. 06. 20

(21) 申请号 202210479781.1

G06V 10/774 (2022.01)

(22) 申请日 2022.05.05

G06V 10/82 (2022.01)

G06N 3/0464 (2023.01)

(65) 同一申请的已公布的文献号

申请公布号 CN 114943859 A

(56) 对比文件

(43) 申请公布日 2022.08.26

CN 113537305 A, 2021.10.22

CN 113655479 A, 2021.11.16

(73) 专利权人 兰州理工大学

CN 114067160 A, 2022.02.18

地址 730050 甘肃省兰州市七里河区兰工
坪路287号

CN 109919183 A, 2019.06.21

CN 109961089 A, 2019.07.02

(72) 发明人 李晓旭 杨世丞 刘俊 燕锦涛
安文娟 张文斌 李睿凡 马占宇
陶剑

CN 110020682 A, 2019.07.16

CN 112288013 A, 2021.01.29

CN 112836773 A, 2021.05.25

(74) 专利代理机构 北京挺立专利事务所(普通
合伙) 11265

CN 113255701 A, 2021.08.13

CN 113723562 A, 2021.11.30

专利代理师 高福勇

CN 113963165 A, 2022.01.21

审查员 孙麒

(51) Int.Cl.

G06V 10/764 (2022.01)

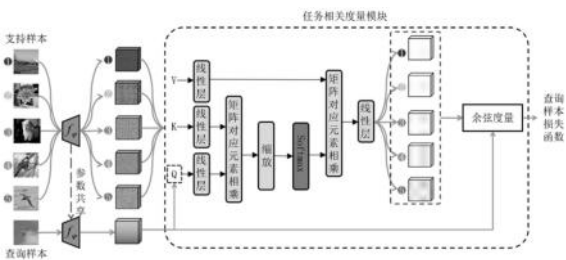
权利要求书3页 说明书9页 附图3页

(54) 发明名称

面向小样本图像分类的任务相关度量学习
方法及装置

(57) 摘要

本发明公开了一种面向小样本图像分类的任务相关度量学习方法及装置,方法主要由数据预处理阶段、构建网络模型阶段、训练模型参数阶段和测试模型性能阶段组成,本发明通过考虑不同任务之间的差异性,引入注意力机制的思想,并学习任务相关的空间映射,利用任务自适应度量学习的方式,解决了小样本图像分类中存在的自适应度量学习问题,从而提高在小样本条件下目标任务分类的准确性,改善了图像的分类效果,具有很高的实用价值。



1. 一种面向小样本图像分类的任务相关度量学习方法,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 D_{train} 和测试集 D_{test} ,训练集 D_{train} 和测试集 D_{test} 的类别空间互斥;

S2、构建面向小样本图像分类的任务相关度量学习模型,模型由嵌入模块 f_ϕ 和任务相关度量模块组成;其中,嵌入模块包含四个卷积块,每个卷积块均包括卷积层、池化层以及非线性激活函数;任务相关度量模块由注意力模块和余弦度量模块组成;

S3、将训练集数据送入面向小样本图像分类的任务相关度量学习模型进行训练,求解模型参数;

步骤S3具体包括:

S301、对于 D_{train} 中的一个任务 T_i ,首先将所有支持样本和查询样本输入嵌入模块 f_ϕ 中;

S302、利用嵌入模块中的卷积神经网络,将支持样本依次经过卷积层、池化层和激活层,最终提取图像的特征 $f_\phi(x_i)$;

S303、将支持样本特征 $F_s \in R^{H \times W \times C}$ 分别作为V和K输入到任务相关度量模块中;

S304、将查询样本中特征 $F_q \in R^{H \times W \times C}$,将其作为Q输入到任务相关度量模块中,其中H和W代表特征空间的大小,C代表特征的通道数;

S305、将V,K,Q分别经过三个权重不同的线性层 $\{W^v, W^k, W^q\} \subset R^{C \times C}$,将提取出来的特征投影到低维,得到转换后特征,表示为 $\{F_s^v, F_s^k, F_q^q\} \subset R^{H \times W \times C}$,公式如下:

$$[F_s^v, F_s^k, F_q^q] = [F_s W^v, F_s W^k, F_q W^q] \quad (1)$$

在公式(1)中, F_s 代表支持样本特征, F_q 代表查询样本特征, W^v, W^k, W^q 代表三个权重不同的线性层, F_s^v, F_s^k, F_q^q 分别代表 F_s 经过 W^v 、 F_s 经过 W^k 、 F_q 经过 W^q 所得到的转化后的特征;公式(1)表示将V,K,Q经过 W^v, W^k, W^q 三个权重不同的线性层投影到低维;

S306、利用公式(2)计算所有支持样本的预测概率,公式如下:

$$F_a = \text{Att}(F_s^v, F_s^k, F_q^q) = \text{softmax}\left(\frac{F_q^q \circ F_s^k}{\sqrt{C}}\right) \circ F_s^v \quad (2)$$

在公式(2)中, \circ 代表矩阵的对应元素相乘, F_s^v, F_s^k, F_q^q 代表经过公式(1)转化后的特征,C代表特征的通道数,softmax代表softmax激活函数, F_a 代表经过公式(2)后得到的加权特征;公式(2)表示求得加权注意力权重后的特征;

S307、将 F_a 再经过一个线性层后得到任务自适应的支持样本特征 $F_A \in R^{H \times W \times C}$;

S308、将查询样本特征Q和任务自适应的支持样本特征 F_A 共同输入到余弦度量模块中,度量模块采用余弦分类器,用于查询样本的分类;余弦度量模块运算公式如下:

$$F = \frac{F_A \circ F_q}{\|F_A\| \|F_q\|} \quad (3)$$

在公式(3)中, F_q 代表查询样本特征, F_A 代表任务自适应的支持样本特征, \circ 代表矩阵的对应元素相乘, $\|A\|$ 表示求矩阵A的二范数,F代表求得的余弦相似度矩阵,公式(3)表示求出任务自适应的支持样本特征 F_A 和查询样本特征 F_q 之间的余弦相似度矩阵;

S309、使用交叉熵损失函数计算支持样本与查询样本的分类预测损失 l_0 ，将 l_0 作为整个网络的总损失loss；交叉熵损失函数公式如下：

$$loss = l_0 = \sum_{i=1}^n y_i \log(p_i) \quad (4)$$

在公式(4)中，n代表种类数量，y代表类标签，若类别是i，则 $y_i=1$ 且其他位为0， p_i 代表类别是i的概率，其值为公式(3)算出来的F矩阵中的对应位置的元素值，loss代表计算后得到的当前损失值，公式(4)表示根据交叉熵损失函数，计算当前的网络模型情况下的损失值；

S310、根据求得的loss使用mini-batch和Adam优化器更新嵌入模块 f_ϕ 和任务相关度量模型的可学习参数，重复训练多个任务，直到网络收敛；

S4、利用训练后的面向小样本图像分类的任务相关度量学习模型对新类任务进行预测，测评模型的性能。

2. 根据权利要求1所述的面向小样本图像分类的任务相关度量学习方法，其特征在于，步骤S1的预处理方法为：从训练集 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 ，同样对于测试集 D_{test} 也有任务 T_k 。

3. 根据权利要求1所述的面向小样本图像分类的任务相关度量学习方法，其特征在于，步骤S2中，每个卷积块包含一个带有64个滤波器的 3×3 的卷积，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，全连接层共128维。

4. 根据权利要求1所述的面向小样本图像分类的任务相关度量学习方法，其特征在于，S310中使用的Adam学习率自适应优化算法具体步骤如下：

S3101、对数据进行初始化：

$$v_{dw}=0, S_{dw}=0, v_{db}=0, S_{db}=0$$

W代表 W_1, W_2, \dots, W_n 的集合，b代表 b_1, b_2, \dots, b_n 的集合，在第t次迭代中，用当前的mini-batch计算W和b的微分dw, db；

S3102、Momentum算法：

根据公式(5)和(6)计算梯度微分的指数加权平均数：

$$v_{dw} = \beta_1 v_{dw} + (1 - \beta_1) dw \quad (5)$$

$$v_{db} = \beta_1 v_{db} + (1 - \beta_1) db \quad (6)$$

S3103、RMSprop算法：

根据公式(7)和(8)计算梯度微分平方的指数加权平均数：

$$S_{dw} = \beta_2 S_{dw} + (1 - \beta_2) (dw)^2 \quad (7)$$

$$S_{db} = \beta_2 S_{db} + (1 - \beta_2) (db)^2 \quad (8)$$

S3104、对两种算法进行偏差修正：

根据公式(9)和(10)进行Momentum算法偏差修正：

$$v_{dw}^{corrected} = \frac{v_{dw}}{1 - \beta_1^t} \quad (9)$$

$$v_{db}^{corrected} = \frac{v_{db}}{1 - \beta_1^t} \quad (10)$$

根据公式(11)和(12)进行RMSprop算法偏差修正；

$$S_{dW}^{corrected} = \frac{S_{dW}}{1 - \beta_2^t} \quad (11)$$

$$S_{db}^{corrected} = \frac{S_{db}}{1 - \beta_2^t} \quad (12)$$

S3105、根据公式(13)和(14)进行梯度下降，更新参数：

$$w := w - \frac{\alpha v_{dW}^{corrected}}{\sqrt{S_{dW}^{corrected}} + \varepsilon} \quad (13)$$

$$b := b - \frac{\alpha v_{db}^{corrected}}{\sqrt{S_{db}^{corrected}} + \varepsilon} \quad (14)$$

在公式(5) - (14)中， v_{dW} , v_{db} , S_{dW} , S_{db} 分别代表有偏差的一阶和二阶矩估计， t 代表次数， α 代表学习率， ε 代表用于数值稳定的小常数， β_1 , β_2 代表矩估计的指数衰减率， dW , db 分别代表 W 和 b 的微分， $v_{dW}^{corrected}$, $v_{db}^{corrected}$, $S_{dW}^{corrected}$, $S_{db}^{corrected}$ 代表经过偏差修正后的一阶和二阶矩估计。

5. 根据权利要求1所述的面向小样本图像分类的任务相关度量学习方法，其特征在于，步骤S4的具体步骤为：

S401、将新类的任务输入训练好的嵌入模块 f_ϕ 中；

S402、将嵌入模块输出的矩阵特征经过任务相关度量模块，得到查询样本与支持样本各类别的余弦度量；

S403、将相似度最高的类作为预测标签，根据预测结果评估模型性能。

6. 一种面向小样本图像分类的任务相关度量学习装置，其特征在于，用以实现权利要求1-5任一项所述的面向小样本图像分类的任务相关度量学习方法，包括以下模块：

数据预处理模块：用于对数据进行预处理，将数据划分成为训练集和测试集并且确定模型的训练方式；

网络模型构建模块：用于引入注意力机制和自适应度量学习，构建面向小样本图像分类的任务相关度量学习模型，模型由嵌入模块 f_ϕ 和任务相关度量模块组成；其中，嵌入模块包含四个卷积块，每个卷积块均包括卷积层、池化层以及非线性激活函数；任务相关度量模块由注意力模块和余弦度量模块组成；

训练模型参数模块：用于面向小样本图像分类的任务相关度量学习模型进行训练，求解模型参数；

测试模型性能模块：利用训练好的面向小样本图像分类的任务相关度量学习模型对新类的任务进行预测，测评模型的性能。

面向小样本图像分类的任务相关度量学习方法及装置

技术领域

[0001] 本发明涉及计算机视觉领域中图像分类,尤其涉及一种面向小样本图像类内共性特征的任务相关度量学习方法及装置。

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] 人类具体通过极少量样本识别一个新物体的能力,例如小朋友只需要看过书中的个别图片,就可以准确的判断什么是“香蕉”或者是“草莓”。小样本学习指的是研究人员希望机器学习模型在学习一定类别的大量数据后,遇到新的类别后,只需要少量的数据就可以快速的学习,实现“小样本学习”。

[0004] 小样本分类属于小样本学习范畴,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。

[0005] 在小样本图像分类的研究方法中,基于深度度量的方法简单而且高效,主要通过比较样本间或者样本与类原型间的距离来判断类别。常常结合数据增强、迁移学习等技术来弥补数据量不足以及模型容易过拟合的缺陷,在很多小样本分类任务上获得了较好的分类性能。但与大样图像分类相比,现有小样本图像分类的性能仍不尽人意,很大程度上限制了小样本图像分类技术的实用化,在自适应的度量学习中还面临着以下问题亟待解决:

[0006] 现有小样本分类方法中,大多假设小样本分类任务使用一个单一的度量方式,例如余弦距离、欧氏距离或一个可学习的度量网络模块。不同的任务包含不同的类别,有些任务适用余弦距离,有些任务适用欧氏距离。因此,如何构建任务自适应的度量也是小样本图像分类值得研究的问题。

发明内容

[0007] 本发明针对上述技术问题,提出一种面向小样本图像分类的任务相关度量学习方法及装置,引入了注意力机制的思想,利用任务自适应度量学习的方式,通过考虑不同任务之间的差异性,并学习任务相关的空间映射,解决了小样本图像分类中存在任务自适应的度量问题,对于图像的分类效果十分明显,具有很高的实用价值。

[0008] 为了实现上述目的,本发明提供如下技术方案:

[0009] 一方面,本发明提供了一种面向小样本图像分类的任务相关度量学习方法,包括以下步骤:

[0010] S1、对数据进行预处理,其中数据包括训练集 D_{train} 和测试集 D_{test} ,训练集 D_{train} 和测试集 D_{test} 的类别空间互斥;

[0011] S2、构建面向小样本图像分类的任务相关度量学习模型,模型由嵌入模块 f_ϕ 和任务相关度量模块组成;其中,嵌入模块包含四个卷积块,每个卷积块均包括卷积层、池化层以及非线性激活函数;任务相关度量模块由注意力模块和余弦度量模块组成;

[0012] S3、将训练集数据送入面向小样本图像分类的任务相关度量学习模型进行训练,求解模型参数;

[0013] S4、利用训练后的面向小样本图像分类的任务相关度量学习模型对新类任务进行预测,测评模型的性能。

[0014] 进一步地,步骤S1的预处理方法为:从训练集 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 ,同样对于测试集 D_{test} 也有任务 T_k 。

[0015] 进一步地,步骤S2中,每个卷积块包含一个带有64个滤波器的 3×3 的卷积,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,全连接层共128维。

[0016] 进一步地,步骤S3具体包括:

[0017] S301、对于 D_{train} 中的一个任务 T_i ,首先将所有支持样本和查询样本输入嵌入模块 f_ϕ 中;

[0018] S302、利用嵌入模块中的卷积神经网络,将支持样本依次经过卷积层、池化层和激活层,最终提取图像的特征 $f_\phi(x_i)$;

[0019] S303、将支持样本特征 $F_s \in R^{HW \times C}$ 分别作为V和K输入到任务相关度量模块中;

[0020] S304、将查询样本中特征 $F_q \in R^{HW \times C}$,将其作为Q输入到任务相关度量模块中,其中H和W代表特征空间的大小,C代表特征的通道数;

[0021] S305、将V,K,Q分别经过三个权重不同的线性层 $\{W^v, W^k, W^q\} \subset R^{C \times C}$,将提取出来的特征投影到低维,得到转换后特征,表示为 $\{F_s^v, F_s^k, F_q^q\} \subset R^{HW \times C}$,公式如下:

$$[0022] \quad [F_s^v, F_s^k, F_q^q] = [F_s W^v, F_s W^k, F_q W^q] \quad (1)$$

[0023] 在公式(1)中, F_s 代表支持样本特征, F_q 代表查询样本特征, W^v, W^k, W^q 代表三个权重不同的线性层, F_s^v, F_s^k, F_q^q 分别代表 F_s 经过 W^v 、 F_s 经过 W^k 、 F_q 经过 W^q 所得到的转化后的特征;公式(1)表示将V,K,Q经过 W^v, W^k, W^q 三个权重不同的线性层投影到低维;

[0024] S306、利用公式(2)计算所有支持样本的预测概率,公式如下:

$$[0025] \quad F_a = \text{Att}(F_s^v, F_s^k, F_q^q) = \text{softmax}\left(\frac{F_q^q \circ F_s^k}{\sqrt{C}}\right) \circ F_s^v \quad (2)$$

[0026] 在公式(2)中, \circ 代表矩阵的对应元素相乘, F_s^v, F_s^k, F_q^q 代表经过公式(1)转化后的特征,C代表特征的通道数,softmax代表softmax激活函数, F_a 代表经过公式(2)后得到的加权特征;公式(2)表示求得加权注意力权重后的特征;

[0027] S307、将 F_a 再经过一个线性层后得到任务自适应的支持样本特征 $F_A \in R^{HW \times C}$;

[0028] S308、将查询样本特征Q和任务自适应的支持样本特征 F_A 共同输入到余弦度量模

块中,度量模块采用余弦分类器,用于查询样本的分类;

[0029] S309、使用交叉熵损失函数计算支持样本与查询样本的分类预测损失 l_0 ,将 l_0 作为整个网络的总损失loss;

[0030] S310、根据求得的loss使用mini-batch和Adam优化器更新嵌入模块 f_φ 和任务相关度量模型的可学习参数,重复训练多个任务,直到网络收敛。

[0031] 进一步地,步骤S308中余弦度量模块运算公式如下:

$$[0032] \quad F = \frac{F_A \circ F_q}{\|F_A\| \|F_q\|} \quad (3)$$

[0033] 在公式(3)中, F_q 代表查询样本特征, F_A 代表任务自适应的支持样本特征, \circ 代表矩阵的对应元素相乘, $\|A\|$ 表示求矩阵A的二范数,F代表求得的余弦相似度矩阵,公式(3)表示求出任务自适应的支持样本特征 F_A 和查询样本特征 F_q 之间的余弦相似度矩阵。

[0034] 进一步地,步骤S309中的交叉熵损失函数公式如下:

$$[0035] \quad loss = l_0 = \sum_{i=1}^n y_i \log(p_i) \quad (4)$$

[0036] 在公式(4)中,n代表种类数量,y代表类标签,若类别是i,则 $y_i=1$ 且其他位为0, p_i 代表类别是i的概率,其值为公式(3)算出来的F矩阵中的对应位置的元素值,loss代表计算后得到的当前损失值,公式(4)表示根据交叉熵损失函数,计算当前的网络模型情况下的损失值。

[0037] 进一步地,S310中使用的Adam学习率自适应优化算法具体步骤如下:

[0038] S3101、对数据进行初始化:

[0039] $v_{dw}=0, S_{dw}=0, v_{db}=0, S_{db}=0$

[0040] W代表 W_1, W_2, \dots, W_n 的集合,b代表 b_1, b_2, \dots, b_n 的集合,在第t次迭代中,用当前的mini-batch计算W和b的微分dW,db;

[0041] S3102、Momentum算法:

[0042] 根据公式(5)和(6)计算梯度微分的指数加权平均数:

$$[0043] \quad v_{dw} = \beta_1 v_{dw} + (1 - \beta_1) dW \quad (5)$$

$$[0044] \quad v_{db} = \beta_1 v_{db} + (1 - \beta_1) db \quad (6)$$

[0045] S3103、RMSprop算法:

[0046] 根据公式(7)和(8)计算梯度微分平方的指数加权平均数:

$$[0047] \quad S_{dw} = \beta_2 S_{dw} + (1 - \beta_2) (dW)^2 \quad (7)$$

$$[0048] \quad S_{db} = \beta_2 S_{db} + (1 - \beta_2) (db)^2 \quad (8)$$

[0049] S3104、对两种算法进行偏差修正:

[0050] 根据公式(9)和(10)进行Momentum算法偏差修正:

$$[0051] \quad v_{dw}^{corrected} = \frac{v_{dw}}{1 - \beta_1^t} \quad (9)$$

$$[0052] \quad v_{db}^{corrected} = \frac{v_{db}}{1 - \beta_1^t} \quad (10)$$

[0053] 根据公式(11)和(12)进行RMSprop算法偏差修正:

$$[0054] \quad S_{dW}^{corrected} = \frac{S_{dW}}{1 - \beta_2^t} \quad (11)$$

$$[0055] \quad S_{db}^{corrected} = \frac{S_{db}}{1 - \beta_2^t} \quad (12)$$

[0056] S3105、根据公式(13)和(14)进行梯度下降,更新参数:

$$[0057] \quad w := w - \frac{\alpha v_{dW}^{corrected}}{\sqrt{S_{dW}^{corrected}} + \varepsilon} \quad (13)$$

$$[0058] \quad b := b - \frac{\alpha v_{db}^{corrected}}{\sqrt{S_{db}^{corrected}} + \varepsilon} \quad (14)$$

[0059] 在公式(5)-(14)中, $v_{dW}, v_{db}, S_{dW}, S_{db}$ 分别代表有偏差的一阶和二阶矩估计, t 代表次数, α 代表学习率, ε 代表用于数值稳定的小常数, β_1, β_2 代表矩估计的指数衰减率, dW, db 分别代表 W 和 b 的微分, $v_{dW}^{corrected}, v_{db}^{corrected}, S_{dW}^{corrected}, S_{db}^{corrected}$ 代表经过偏差修正后的一阶和二阶矩估计。

[0060] 进一步地,步骤S4的具体步骤为:

[0061] S401、将新类的任务输入训练好的嵌入模块 f_ϕ 中;

[0062] S402、将嵌入模块输出的矩阵特征经过任务相关度量模块,得到查询样本与支持样本各类别的余弦度量;

[0063] S403、将相似度最高的类作为预测标签,根据预测结果评估模型性能。

[0064] 另一方面,本发明还提供了一种面向小样本图像分类的任务相关度量学习装置,用以实现上述的任一项方法,包括以下模块:

[0065] 数据预处理模块:用于对数据进行预处理,将数据划分成为训练集和测试集并且确定模型的训练方式;

[0066] 网络模型构建模块:用于引入注意力机制和自适应度量学习,构建面向小样本图像分类的任务相关度量学习模型,模型由嵌入模块 f_ϕ 和任务相关度量模块组成;其中,嵌入模块包含四个卷积块,每个卷积块均包括卷积层、池化层以及非线性激活函数;任务相关度量模块由注意力模块和余弦度量模块组成;

[0067] 训练模型参数模块:用于面向小样本图像分类的任务相关度量学习模型进行训练,求解模型参数;

[0068] 测试模型性能模块:利用训练好的面向小样本图像分类的任务相关度量学习模型对新类的任务进行预测,测评模型的性能。

[0069] 与现有技术相比,本发明的有益效果为:

[0070] 本发明引入注意力机制的思想,建立了一种面向小样本图像分类的任务相关度量学习方法及装置,通过考虑不同任务之间的差异性,并学习任务相关的空间映射,利用任务自适应度量学习的方式,解决了小样本图像分类中存在的自适应度量学习问题,从而提高在小样本条件下目标任务分类的准确性,改善了图像的分类效果,具有很高的实用价值。

附图说明

[0071] 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。

[0072] 图1为本发明实施例提供的面向小样本图像分类的任务相关度量学习方法的阶段流程图。

[0073] 图2为本发明实施例提供的面向小样本图像分类的任务相关度量学习模型结构图。

[0074] 图3为本发明实施例提供的面向小样本图像分类的任务相关度量学习模型的功能模块构成图。

[0075] 图4为本发明实施例提供的嵌入模块 f_{ϕ} 结构图。

[0076] 图5为本发明实施例提供的特征矩阵经过三个权重不同的线性层示意图。

[0077] 图6为本发明实施例提供的注意力机制运算示意图。

具体实施方式

[0078] 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。

[0079] 根据本文公开的一个方面,提供了一种面向小样本图像分类的任务相关度量学习方法,如图1所示,包括以下阶段步骤:

[0080] S1、数据预处理阶段:对数据进行预处理,其中数据包括训练集和测试集;

[0081] S2、构建网络模型阶段:引入注意力机制和自适应度量学习,构建面向小样本图像分类的任务相关度量学习模型;

[0082] S3、训练模型参数阶段:将训练集数据送入面向小样本图像分类的任务相关度量学习模型进行训练,求解模型参数;

[0083] S4、测试模型性能阶段:利用训练后的面向小样本图像分类的任务相关度量学习模型对新类任务进行预测,测评模型的性能。

[0084] 在一些实施例中,所述阶段步骤S1包括以下子步骤:

[0085] S101、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 $D_{test} = \{(x_i^*, y_i^*), y_i^* \in L_{test}\}_{i=1}^{N_{test}}$ 两个部分,且这两个部分的类别空间互斥。将 D_{train} 作为基类数据训练模型, D_{test} 作为新类数据测评模型性能;

[0086] S102、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 ,同样对于 D_{test} 也有任务 T_k 。

[0087] 在一些实施例中,所述步骤S2包括以下步骤:

[0088] 构建面向小样本图像分类的任务相关度量学习模型,其结构如图2所示,网络模型

划分为嵌入式模块和任务相关度量模块,如图3所示。嵌入式模块由输入层、卷积层、池化层以及激活函数组成,如图4所示,目的是为了提取样本的局部特征;其中,在遵循四层卷积架构来形成特征提取器 f_φ ,每个块包含一个带有64个滤波器的 3×3 的卷积,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,全连接层共128维。

[0089] 任务相关度量模块由注意力模块和余弦度量模块组成,目的是通过考虑不同任务之间的差异性,并学习任务相关的空间映射,从而提高在小样本条件下目标任务分类的准确性。

[0090] 在一些实施例中,所述步骤S3包括以下子步骤:

[0091] S301、对于 D_{train} 中的一个任务 T_i ,首先将所有支持样本和查询样本输入嵌入模块 f_φ 中;

[0092] S302、利用嵌入模块中的卷积神经网络,将支持样本依次经过卷积层、池化层和激活层,最终提取图像的特征 $f_\varphi(x_i)$;

[0093] S303、将支持样本特征 $F_s \in R^{HW \times C}$ 分别作为V和K输入到任务相关度量模块中;

[0094] S304、将查询样本中特征 $F_q \in R^{HW \times C}$,将其作为Q输入到任务相关度量模块中,其中H和W代表特征空间的大小,C代表特征的通道数;

[0095] S305、将V,K,Q分别经过三个权重不同的线性层 $\{W^v, W^k, W^q\} \subset R^{C \times C}$,将提取出来的特征投影到低维,得到转换后特征,表示为 $\{F_s^v, F_s^k, F_q^q\} \subset R^{HW \times C}$,公式如下:

$$[0096] \quad [F_s^v, F_s^k, F_q^q] = [F_s W^v, F_s W^k, F_q W^q] \quad (1)$$

[0097] 在公式(1)中, F_s 代表支持样本特征, F_q 代表查询样本特征, W^v, W^k, W^q 代表三个权重不同的线性层, F_s^v, F_s^k, F_q^q 分别代表 F_s 经过 W^v 、 F_s 经过 W^k 、 F_q 经过 W^q 所得到的转化后的特征;公式(1)表示将V,K,Q经过 W^v, W^k, W^q 三个权重不同的线性层投影到低维,如图5。

[0098] S306、利用公式(2)计算所有支持样本的预测概率,公式如下:

$$[0099] \quad F_a = \text{Att}(F_s^v, F_s^k, F_q^q) = \text{softmax}\left(\frac{F_q^q \circ F_s^k}{\sqrt{C}}\right) \circ F_s^v \quad (2)$$

[0100] 在公式(2)中, \circ 代表矩阵的对应元素相乘, F_s^v, F_s^k, F_q^q 代表经过公式(1)转化后的特征,C代表特征的通道数,softmax代表softmax激活函数, F_a 代表经过公式(2)后得到的加权特征,公式(2)表示求得加权注意力权重后的特征,如图6所示。

[0101] 在公式(2)中,对经过运算后的 $\frac{F_q^q \circ F_s^k}{\sqrt{C}}$ 矩阵使用softmax激活函数, $F_q^q \circ F_s^k$ 表示两个特征矩阵的对应元素相乘,即是将 F_s^k 转置后与 F_q^q 进行矩阵相乘,softmax激活函数分别对每一行、每一列进行归一化得到注意力权重矩阵 $\alpha \in R^{HW \times HW}$,使用这种方法可以保持梯度的稳定。

[0102] S307、将 F_a 再经过一个线性层后得到任务自适应的支持样本特征 $F_A \in R^{HW \times C}$ 。

[0103] S308、接着将查询样本特征Q和任务自适应的支持样本特征 F_A 共同输入到余弦度

量模块中,度量模块采用余弦分类器,用于查询样本的分类。

[0104] 采用的余弦度量模块,其运算公式如下:

$$F = \frac{F_A \circ F_q}{\|F_A\| \|F_q\|} \quad (3)$$

[0106] 在公式(3)中, F_q 代表查询样本特征, F_A 代表任务自适应的支持样本特征, \circ 代表矩阵的对应元素相乘, $\|A\|$ 表示求矩阵A的二范数,F代表求得的余弦相似度矩阵。公式(3)表示求出任务自适应的支持样本特征 F_A 和查询样本特征 F_q 之间的余弦相似度矩阵。

[0107] S309、使用交叉熵损失函数计算支持样本与查询样本的分类预测损失 l_0 ,将 l_0 作为整个网络的总损失loss。

[0108] 交叉熵损失函数公式如下:

$$loss = l_0 = \sum_{i=1}^n y_i \log(p_i) \quad (4)$$

[0110] 在公式(4)中,n代表种类数量,y代表类标签,若类别是i,则 $y_i=1$ 且其他位为0, p_i 代表类别是i的概率,其值为公式(3)算出来的F矩阵中的对应位置的元素值,loss代表计算后得到的当前损失值。公式(4)表示根据交叉熵损失函数,计算当前的网络模型情况下的损失值。

[0111] S310、根据求得的loss使用mini-batch和Adam优化器更新嵌入模块 f_ϕ 和任务相关度量模型的可学习参数,重复训练多个任务,直到网络收敛。

[0112] 使用的Adam学习率自适应优化算法具体步骤如下:

[0113] S3101、为了简化描述,W代表 W_1, W_2, \dots, W_n 的集合,b代表 b_1, b_2, \dots, b_n 的集合。

[0114] 对数据进行初始化: $v_{dw}=0, S_{dw}=0, v_{db}=0, S_{db}=0$ 。

[0115] 在第t次迭代中,用当前的mini-batch计算W和b的微分dW,db。

[0116] S3102、Momentum算法

[0117] 公式计算梯度微分的指数加权平均数:

$$v_{dw} = \beta_1 v_{dw} + (1 - \beta_1) dW \quad (5)$$

$$v_{db} = \beta_1 v_{db} + (1 - \beta_1) db \quad (6)$$

[0120] S3103、RMSprop算法公式计算梯度微分平方的指数加权平均数:

$$S_{dw} = \beta_2 S_{dw} + (1 - \beta_2) (dW)^2 \quad (7)$$

$$S_{db} = \beta_2 S_{db} + (1 - \beta_2) (db)^2 \quad (8)$$

[0123] S3104、对两种算法都进行偏差修正:

[0124] 1) Momentum算法偏差修正:

$$v_{dw}^{corrected} = \frac{v_{dw}}{1 - \beta_1^t} \quad (9)$$

$$v_{db}^{corrected} = \frac{v_{db}}{1 - \beta_1^t} \quad (10)$$

[0127] 2) RMSprop算法偏差修正:

$$[0128] \quad S_{dW}^{corrected} = \frac{S_{dW}}{1 - \beta_2^t} \quad (11)$$

$$[0129] \quad S_{db}^{corrected} = \frac{S_{db}}{1 - \beta_2^t} \quad (12)$$

[0130] S3105、进行梯度下降,更新参数:

$$[0131] \quad w := w - \frac{\alpha v_{dW}^{corrected}}{\sqrt{S_{dW}^{corrected}} + \varepsilon} \quad (13)$$

$$[0132] \quad b := b - \frac{\alpha v_{db}^{corrected}}{\sqrt{S_{db}^{corrected}} + \varepsilon} \quad (14)$$

[0133] 在公式(5) - (14)中, v_{dW} , v_{db} , S_{dW} , S_{db} 分别代表有偏差的一阶和二阶矩估计, t 代表次数, α 代表学习率, ε 代表用于数值稳定的小常数, β_1 , β_2 代表矩估计的指数衰减率, dW , db 分别代表 W 和 b 的微分, $v_{dW}^{corrected}$, $v_{db}^{corrected}$, $S_{dW}^{corrected}$, $S_{db}^{corrected}$ 代表经过偏差修正后的一阶和二阶矩估计。

[0134] Mini-batch算法优点:把数据分为若干个批,按批来更新参数,这样,一个批中的一组数据共同决定了本次梯度的方向,下降起来就不容易跑偏,减少了随机性。另一方面因为批的样本数与整个数据集相比小了很多,计算量也不是很大,更加便于计算。

[0135] Adam优化算法优点:动量直接并入了梯度一阶矩(指数加权)的估计,将动量应用于缩放后的梯度。包括了偏差修正步骤,修正从原点初始化的一阶矩(动量项)和(非中心的)二阶矩估计。

[0136] 在一些实施例中,所述步骤S4包括以下子步骤:

[0137] S401、将新类的任务输入训练好的嵌入模块 f_ϕ 中;

[0138] S402、将嵌入模块输出的矩阵特征经过任务相关度量模块,得到查询样本与支持样本各类别的余弦度量;

[0139] S403、将相似度最高的类作为预测标签,根据预测结果评估模型性能。

[0140] 根据本文公开的另一个方面,本发明还提供了一种面向小样本图像分类的任务相关度量学习装置,用于实现上述面向小样本图像分类的任务相关度量学习方法,包括:

[0141] 数据预处理模块:对数据进行预处理,将数据划分成为训练集和测试集并且确定模型的训练方式;

[0142] 网络模型构建模块:引入注意力机制和自适应度量学习,构建面向小样本图像分类的任务相关度量学习模型;模型由嵌入模块 f_ϕ 和任务相关度量模块组成;其中,嵌入模块包含四个卷积块,每个卷积块均包括卷积层、池化层以及非线性激活函数;任务相关度量模块由注意力模块和余弦度量模块组成;

[0143] 训练模型参数模块:用面向小样本图像分类的任务相关度量学习模型进行训练,求解模型参数;

[0144] 测试模型性能模块:利用训练好的面向小样本图像分类的任务相关度量学习模型对新类的任务进行预测,测评模型的性能。

[0145] 以上结合附图对所提出的面向小样本图像分类的任务相关度量学习方法及模型

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。

[0146] 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。

[0147] 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。

[0148] 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殊进行等同替换;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

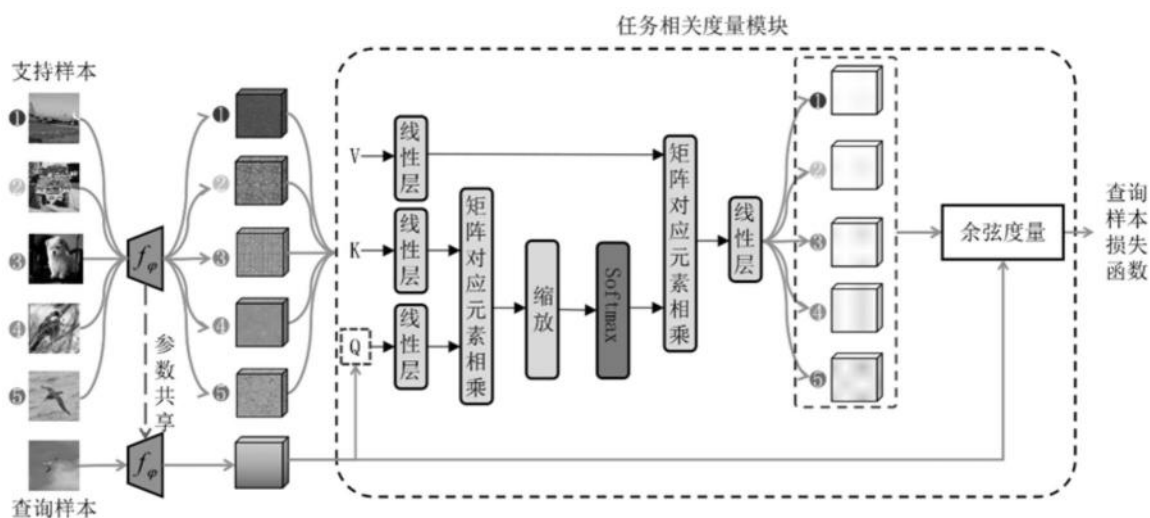


图2

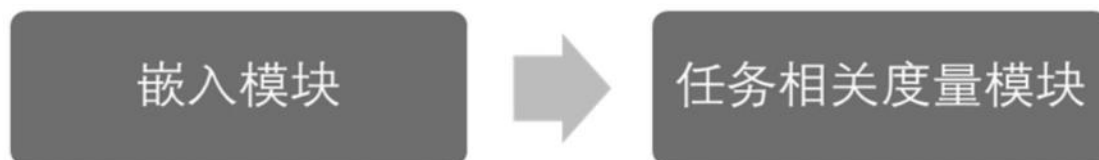


图3

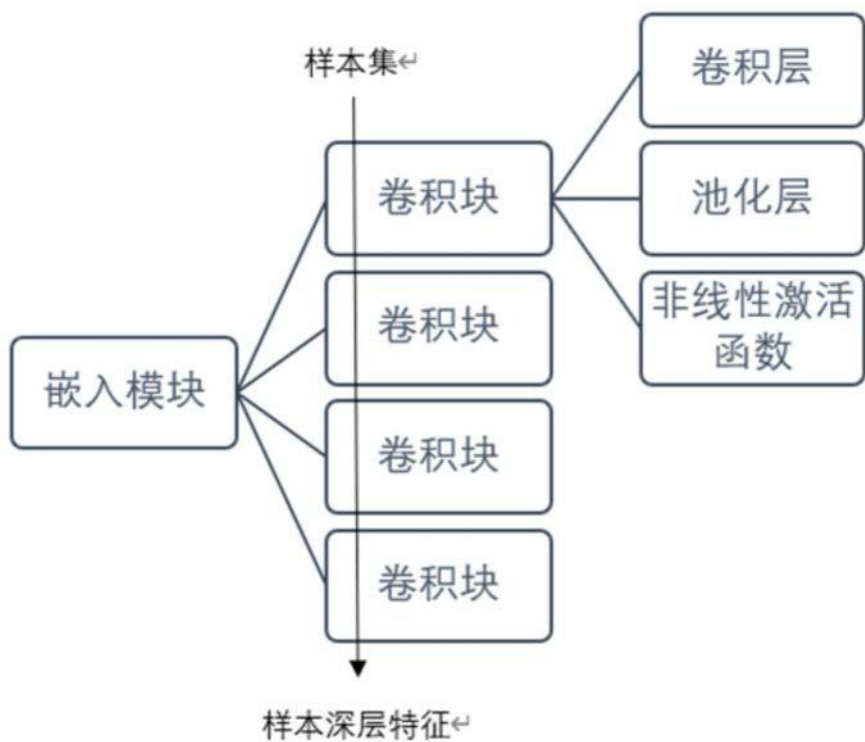


图4

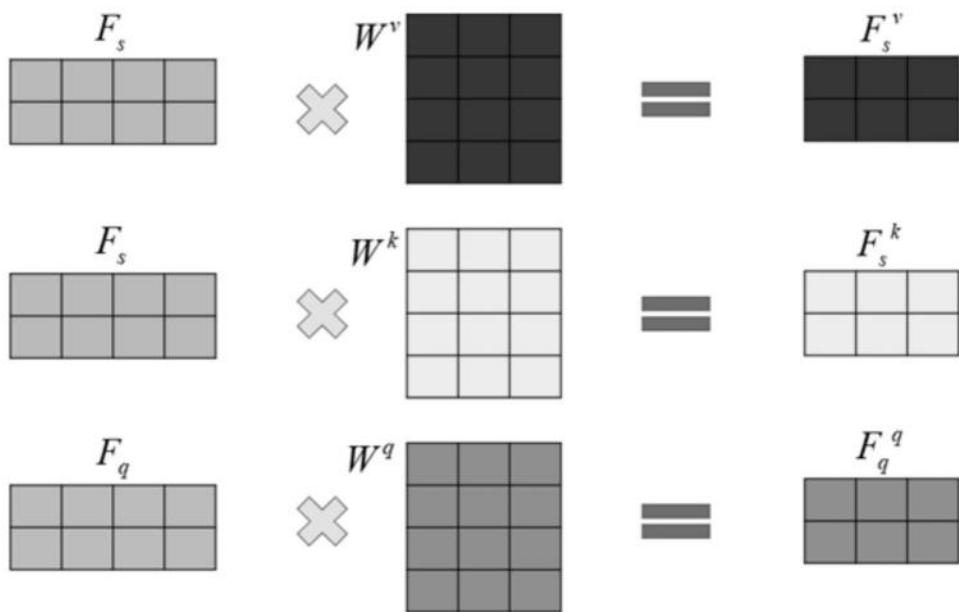


图5

F_q^q

\times

F_s^k

$\frac{\text{softmax}(\text{---})}{\sqrt{C}}$

F_s^v

F_a

图6