

使用模糊聚类的胶囊网络在图像分类上的研究

张天柱 邹承明

(武汉理工大学计算机科学与技术学院 武汉 430070)

摘要 胶囊网络中动态路由的本质就是聚类算法思想的实现。考虑到已有胶囊网络中的聚类方式需要数据满足一定的分布才能达到最好的效果,且图像特征比较复杂,于是将一种普适性更好的模糊聚类算法作为胶囊网络中的特征整合方式,并添加了一个使用信息熵来度量不确定性的激活值,以区分同一级别胶囊层特征的显著性。同时,借鉴特征金字塔网络的思想,将不同胶囊层的特征采样成相同尺度,然后进行融合独立训练。基于 Keras 框架进行实验,其结果表明,相比原来的胶囊网络,这种具有新型结构的胶囊网络在 MNIST 和 CIFAR-10 上有更高的识别准确率。对比实验证明了模糊聚类算法在胶囊网络上的应用潜力,其改善了原胶囊网络中聚类算法存在局限的问题;也证明了胶囊网络中不同层的特征进行融合,能够获得包含信息更丰富且表达能力更强的特征。

关键词 图像分类,胶囊网络,模糊聚类算法,特征金字塔网络,多尺度特征融合

中图分类号 TP391 文献标识码 A DOI 10.11896/jsjcx.190200315

Study on Image Classification of Capsule Network Using Fuzzy Clustering

ZHANG Tian-zhu ZOU Cheng-ming

(School of Computer Science and Technology, Wuhan University of Technology, Wuhan 430070, China)

Abstract The essence of dynamic routing in capsule network is the implementation of clustering algorithm. Considering that the clustering method used in the previous capsule network requires the data to meet certain distributions to achieve the best effect while features of image are complicated, a more universal fuzzy clustering algorithm was taken as the feature integration scheme to replace the old in this paper. And an activation value using information entropy to measure the indeterminacy was added to the model, so as to distinguish the significance of capsule features at the same layer. Meanwhile, drawing on the idea of feature pyramid network, the features of different capsule layers are sampled to the same size to fuse and then are trained independently. Experimental results based on the Keras framework show that the capsule network with new structure has higher recognition accuracy on MNIST and CIFAR-10 than the original capsule network. The contrast experiments prove great potential of fuzzy clustering algorithm applying on capsule network, which alleviates the limitation of the clustering algorithm in the original capsule network. The results also prove that the features of different layers in the capsule network can be fused to be more informative and expressive.

Keywords Image classification, Capsule network, Fuzzy clustering algorithm, Feature pyramid network, Multi-scale feature fusion

1 引言

图像分类是计算机视觉^[1]领域中一个经典的研究课题,目前的主流方法大多是基于卷积神经网络(Convolutional Neural Network, CNN)的工作。Yann 等^[2]于 20 世纪 90 年代确立了卷积神经网络的基本结构,并在此基础上完善了 LeNet 模型,该模型在手写识别任务上取得了巨大进展,此后卷积神经网络便一直被广泛研究。这些关于卷积神经网络的研究一直致力于在能够接受的计算资源内增加网络的深度或宽度,以提高模型在测试集上的准确率^[3-9],或使用各种各样

的技巧尽量避免网络在层数过多时产生过拟合问题^[10-12]。但是,卷积神经网络还是保持着卷积层、池化层和全连接层的基本结构^[13],网络层数过深时易产生过拟合和训练负担大的问题。

于是, Sabour 等^[14]于 2017 年首次提出了胶囊网络。胶囊网络是一种具有全新结构的网络,它将神经元的输入由标量提升为向量,通过动态路由来调整网络参数以对特征进行整合。相比于标量,向量不仅能表达信息的有无,还能表达信息的种类,因此胶囊网络极大地丰富了特征的表达能力。文献^[14]证明了胶囊网络具有分割高度重叠数字的能力。Hin-

到稿日期:2019-02-18 返修日期:2019-06-20 本文受湖北省自然科学基金资助项目(2017CFB302)资助。

张天柱(1993-),男,硕士生,主要研究方向为深度学习、计算机视觉;邹承明(1975-),博士,教授,CCF 会员,主要研究方向为计算机视觉、嵌入式系统等, E-mail: zoucm@whut.edu.cn(通信作者)。

ton等^[15]于2018年提出了使用期望最大化路由(EM Routing)的矩阵胶囊网络,其中的实验结果表明胶囊网络不仅能够很好地学习不同视点的物体变化,而且相比于传统卷积神经网络天生具有更好的泛化能力。胶囊网络中动态路由的本质就是聚类算法思想在模型中的实现。文献[14]中的动态路由过程是基于K-Means聚类算法实现的;文献[15]中则将高斯混合模型(Gaussian Mixture Model, GMM)结合到特征整合过程中,通过EM路由对网络参数进行调整。在文献[14]和文献[15]工作的基础上,Jaiswal等^[16]用胶囊网络代替标准卷积神经网络作为生成对抗网络中的鉴别器;Deng等^[17]使用双层的胶囊网络分别训练简单数据集和复杂数据集;Zhang等^[18]证明了胶囊网络在小规模数据集上的有效性,以及不对训练数据使用增强技术时优于卷积神经网络;Xiang等^[19]提出了一种多尺度的胶囊网络,通过提取多尺度特征来获取结构和语义信息,然后将特征的阶层编码为多维初级胶囊后作为最初的输入。这些使用胶囊网络或基于胶囊网络对模型进行改进的工作都各自取得了不俗的效果,体现了胶囊网络广阔的应用前景。

上述基于胶囊网络的工作仍然使用了文献[14]或文献[15]中的动态路由过程,而我们发现这两项工作中的聚类算法都存在一定的局限性,K-Means聚类算法不适合形状特征太复杂的数据,GMM则是在数据满足正态分布时才能达到最好的效果,而图像特征往往比较复杂,难以很好地满足这些条件。模糊聚类算法FCM^[20]是一种基于模糊数学的柔性划分聚类算法,每个样本以隶属度的形式属于各个聚类中心。由于胶囊网络中一个低级特征通常不仅仅只属于一种高级特征,而模糊聚类算法中以隶属度将样本分配给聚类中心的形式很适用于表达胶囊网络中低层特征和高层特征之间的关系。基于上述原因,本文将模糊聚类算法结合到胶囊网络中,以提升整合特征的表达能力。为了使用模糊聚类算法来实现胶囊网络中的动态路由过程,且衡量某个胶囊特征相比其他同一级别胶囊特征的显著性,本文在原生FCM聚类算法的基础上设计了一种使用信息熵来度量特征的不确定性,以区分同一级别特征显著性的算法SFCM,并将其作为胶囊网络中相邻胶囊层之间对特征进行整合的动态路由迭代方式。

另外,金字塔结构是图像领域中一个常用且有效的结构,在人工提取特征的时代就已被广泛使用。最初,基于图像金字塔构建的金字塔结构特征^[21]主要是为了识别不同尺度的物体,ION^[22]通过RoI Pooling操作从不同层中提取区域特征,HyperNet^[23]使用反卷积将不同层上的特征直接连接起来,这种对图像金字塔的每个级别都进行特征化的方式可以产生多尺度的特征表示,它对于图像分类同样有效。但在利用网络进行特征提取的时代,由于在网络中加入金字塔结构会极大地增加计算量和内存损耗,因此图像金字塔一直被避免使用。Lin等^[24]于2017年提出了特征金字塔网络(Feature Pyramid Networks, FPN),使用自底向上和横向连接的方式将不同尺度的特征融合在一起,具有这种结构的特征金字塔网络使用独立训练的方式,不但没有加剧网络的计算损耗,而

且在很大程度上解决了图像金字塔方法使得网络负担过重的问题。借鉴特征金字塔网络的思想,将这种融合特征并进行独立训练的特征金字塔网络结构应用到胶囊网络中,以进一步提升胶囊网络在图像分类任务上的表现。

本文提出的模型在MNIST和CIFAR-10数据集上取得了比文献[14]和文献[15]中模型更好的表现,证明了将模糊聚类算法结合到胶囊网络中的潜力和将多尺度特征融合并进行独立训练的方式应用在胶囊网络上的有效性。目前,胶囊网络刚被提出不久,其研究尚处于起步状态,就在一些图像分类和语义识别任务上获得了不俗的表现^[25];而胶囊网络中还存在许多值得改进和探索的地方,寻找更有效的胶囊网络新结构是一个值得研究的课题。

2 胶囊网络的核心特性

胶囊网络中包含多个胶囊层,每一个胶囊层都被分成许多组小神经元,称之为胶囊,胶囊的输出是向量。其中,活跃胶囊向量的总长度表示实体存在的概率,向量的方向则表示实例化参数(如姿势、形变、色调、纹理等)。同一级别的活跃胶囊通过动态路由对更高级别胶囊的实例化参数进行预测,当多个预测一致时,则激活更高级别的胶囊。胶囊网络与卷积神经网络的主要区别是:胶囊网络用向量输出替换CNN中输出为标量的特征检测器,并用协议路由替换最大池化。另外,为了保留CNN中跨空间地复制已学到知识的优点,胶囊网络中除了最后胶囊层以外的所有胶囊都保持了卷积结构。与CNN一样,更高级别的胶囊能够覆盖更大的图像区域。然而,与最大池化不同,胶囊网络通过协议路由由这种特征传递方式,能够部分保留区域内实体的精确位置信息。

为了使胶囊网络具有多角度的特征表达能力,每个胶囊的输出在传入下一个胶囊之前,都要乘以姿势矩阵 M_j 进行变换。矩阵 M_j 为权值共享矩阵,即对于固定的上层胶囊 v_j ,它与所有下层胶囊相连接的姿势矩阵是共用的,如图1所示。

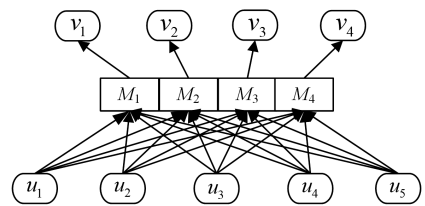


图1 相邻胶囊层之间的特征通过姿势矩阵传递

Fig. 1 Features between adjacent capsule layers are transmitted through pose matrix

该姿势矩阵解决了聚类初始化的问题,同一输入经过矩阵 M_j 可以映射为不同的初始化,而且改变输入向量 u_i 的维度可以改变聚类后的中心向量 v_j 的维度,通过升高或降低向量维度可以保证中心向量 v_j 的特征表达能力。

3 使用模糊聚类的胶囊网络模型结构

由于本文模型涉及的参数较多,为了便于理解,我们增加了各个参数详细释义的统一符号列表,如表1所列。

表 1 统一符号列表
Table 1 Unified symbol list

符号	释义
n	低级胶囊层中胶囊单元的总个数
k	高级胶囊层中胶囊单元的总个数
d	姿势矩阵的维度
m	表示样本轻缓程度的超参数
u_i	低级胶囊层中的第 i 个胶囊单元
v_j	高级胶囊层中的第 j 个胶囊单元
M_j	v_j 与低级胶囊层中所有胶囊单元相联系的姿势矩阵
W_{ij}	u_i 与 v_j 之间相联系的可变换矩阵
Ω_L	L 胶囊层
a_j	胶囊单元 v_j 的激活值
a_i^{last}	低级胶囊层中单元 u_i 的最终激活值
S_j	表示聚类中心 v_j 稳定性的熵值
R_{ij}	样本 u_i 选取某一聚类中心 v_j 的概率
r_{ij}	R_{ij} 的 Softmax 结果
β_α, β_μ	上层胶囊的训练参数
cost_j	激活值的损耗部分,用以简化表示

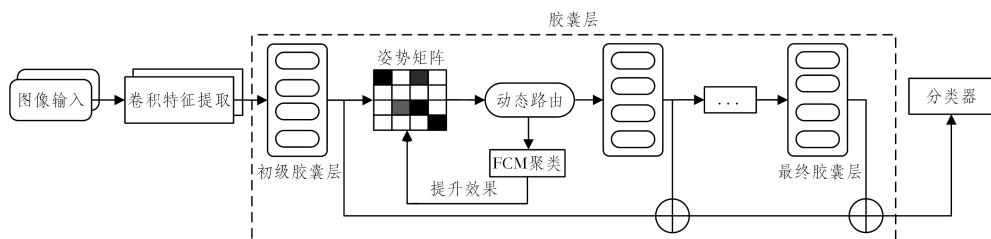


图 2 使用模糊聚类和多尺度特征融合独立训练的胶囊网络模型结构

Fig. 2 Model structure of capsule network using fuzzy clustering and with multi-scale feature fusion independent training

模型中的 L 胶囊层记为 Ω_L , 其中共有 n 个胶囊单元 u_i ($i=1, 2, \dots, n$); 其下一胶囊层记为 Ω_{L+1} , 其中共有 k 个胶囊单元 v_j ($j=1, 2, \dots, k$)。为了方便表述,本文中关于相邻胶囊层之间动态路由过程的说明均以 Ω_L 和 Ω_{L+1} 为对象。 L 胶囊层中的胶囊单元首先经过姿势矩阵 M_j (该姿势矩阵的作用是通过改变矩阵上不同元素的值,经映射后学习到一个特征在不同视点的变化)进行映射初始化, L 胶囊层与 $L+1$ 胶囊层之间有一个相联的 $i \times j$ 维变换矩阵 W_{ij} , 该变换矩阵与模糊聚类算法中的隶属度矩阵对应,且每个 $L+1$ 胶囊层中的胶囊单元 v_j 都有一个相应的激活值 a_j 。姿势矩阵、变换矩阵和激活值是训练过程中被主要训练的参数,其中姿势矩阵通过反向传播进行优化,变换矩阵和激活值在动态路由过程中进行迭代更新。 L 胶囊层中的胶囊单元 u_i 通过姿势矩阵和变换矩阵对下一胶囊层中的胶囊单元 v_j 进行投票,使用一个非线性的路由过程,获得所有 $i \in \Omega_L$ 和 $j \in \Omega_{L+1}$ 的输入 W_{ij} 和 a_j , 该非线性过程迭代地调整 $L+1$ 胶囊层中胶囊单元的平均值、方差和激活概率,以及所有 $i \in \Omega_L$ 和 $j \in \Omega_{L+1}$ 之间的分配概率。

理论上,在图像输入比较复杂时,初次提取的特征也更加复杂,增加胶囊层的个数是增强整合后特征表达能力最直观的方法。由于更高级别的胶囊层中需要包含更丰富的信息,因此,其中胶囊单元的个数应该大于或等于上一胶囊层中单元的个数。结合考虑网络的计算损耗,胶囊层的个数和融合特征训练的层数不应被盲目增加,而应根据具体场景进行选择。

使用模糊聚类的胶囊网络模型结构如图 2 所示,图像输入经过卷积层提取出最初的特征,然后将得到的多组特征图展成一维,按对应元素的位置合并来组成向量表示的输入,并将其输送至初级胶囊层。各个相邻胶囊层之间通过动态路由对特征进行整合,该过程修正下层胶囊与上层胶囊之间的变换矩阵和激活值等参数,以形成最优整合结果。本文模型的动态路由过程基于模糊聚类算法的思想来实现。图 2 中的 \oplus 表示不同胶囊层之间进行特征融合的过程,先将最终胶囊层提取出的特征输送至分类器进行一次训练,然后将最终胶囊层特征上采样成与上一胶囊层特征同样的尺度进行融合,再进行一次训练;针对其他层的特征融合训练也是采用类似的操作。其中,提取特征的卷积层规格、胶囊层的个数、各个胶囊层中胶囊单元的个数和特征融合训练层数均可以根据具体场景进行调整。

3.1 使用信息熵度量不确定性的模糊聚类算法

针对文献[14]和文献[15]中两种聚类算法存在的局限性,文中尝试将一种普适性更好的模糊聚类算法应用到胶囊网络中。模糊聚类算法主要包含以下几个关键步骤,其中本文采用欧氏距离作为相似性的衡量方式。

步骤 1 计算隶属度矩阵 W_{ij} :

$$W_{ij} = \frac{1}{\sum_{c=1}^k \left(\frac{\|u_i - v_c\|}{\|u_i - v_j\|} \right)^{\frac{2}{m-1}}} \quad (1)$$

步骤 2 计算并更新聚类中心:

$$v_j = \frac{\sum_{i=1}^n W_{ij}^m u_i}{\sum_{i=1}^n W_{ij}^m} \quad (2)$$

步骤 3 计算当前目标函数的值:

$$J = \sum_{j=1}^k \sum_{i=1}^n W_{ij}^m \|u_i - v_j\|^2 \quad (3)$$

步骤 4 若最近两次目标函数的差值小于阈值或已达到最大迭代次数,则迭代结束,获取聚类结果;否则返回步骤 1 继续执行。

式(1)—式(3)中, k 为聚类中心的个数,与高级胶囊层中单元的个数相对应; n 为样本中数据的个数,与低级胶囊层中单元的个数相对应; W_{ij} 为样本 u_i (低级胶囊层中的第 i 个单元)属于聚类中心 v_j (高级胶囊层中的第 j 个单元)的隶属度,该隶属度在初级胶囊层被平均初始化; m 是一个表示样本轻缓程度的超参数,过大会导致聚类效果很差,过小则会接近硬聚类算法。文中将 m 初始化为 2,在一次动态路由结束后,令

其自增 1, 然后比较两次目标函数的值, 如果此轮的目标函数值相比上一轮的目标函数值的减小量超过了一个指定阈值, 则继续增大 m 的值。

由于模糊聚类算法是基于加权欧氏距离的聚类, 其聚类中心向量是类内向量的加权平均, 而这种平均无法体现文献 [14] 中以向量模长表示特征显著性 (分支向量越多且越靠近, 则特征越显著) 的特点, 因此我们增加一个标量 a_j 来表示某一胶囊特征相比于其他同一级别胶囊特征的显著性。该标量被称为激活值。用一个聚类中心的不确定性来描述这个激活值, 不确定性越大, 说明该聚类中心越不稳定, 此时激活值越小; 不确定性越小, 则说明分布越集中, 处于稳定状态, 激活值越大。该不确定性用信息熵度量:

$$S_j = - \int p(u_i | v_j) \ln(p(u_i | v_j)) du_i \quad (4)$$

其中, $p(u_i | v_j)$ 表示一个聚类中心中样本的概率分布, 且有:

$$p(u_i) = \sum_{j=1}^k p(v_j) p(u_i | v_j) \quad (5)$$

其中, $p(v_j)$ 表示选取某一聚类中心的概率, 与类内样本 x 无关:

$$p(v_j) = \frac{W_{ij}}{\sum_{j=1}^k W_{ij}} \quad (6)$$

胶囊网络一次路由迭代中的已知参数主要为隶属度矩阵的当前元素值 W_{ij} , 即样本 u_i 属于聚类中心 v_j 的概率 $p(u_i | v_j)$, 而我们需要求取 $p(v_j | u_i)$, 即样本 u_i 属于聚类中心 v_j 的概率。经条件概率公式转换, 得到:

$$\begin{aligned} p(v_j | u_i) &= \frac{p(u_i | v_j) p(v_j)}{p(u_i)} \\ \Rightarrow p(u_i | v_j) &= \frac{p(v_j | u_i) p(u_i)}{p(v_j)} \end{aligned} \quad (7)$$

将式(7)代入式(4), 得到:

$$\begin{aligned} S_j &= - \frac{1}{p(v_j)} \int p(v_j | u_i) p(u_i) \ln(p(v_j | u_i)) du_i \\ &= - \frac{1}{p(v_j)} E[p(v_j | u_i) \ln(p(v_j | u_i))] \end{aligned} \quad (8)$$

根据期望值的计算公式, 并引入迭代过程中当前的激活值 \hat{a}_j , 得到聚类中心的选取概率 R_{ij} :

$$R_{ij} = \hat{p}(v_j) = \frac{\hat{a}_j W_{ij}}{\hat{a}_j \sum_{j=1}^k W_{ij}} \quad (9)$$

将式(9)代入式(8), 得到:

$$S_j = - \frac{1}{\sum_{i=1}^n R_{ij}} \sum_{i=1}^n R_{ij} \ln(W_{ij}) = - \sum_{i=1}^n r_{ij} \ln(W_{ij}) \quad (10)$$

为了简化表示, 令:

$$r_{ij} = \frac{R_{ij}}{\sum_{i=1}^n R_{ij}} \quad (11)$$

由于熵越小时特征越显著, 须对 S_j 取负, 且在神经网络模型中, 出于训练需要和常规认知, 需要将激活值压缩到 0~1 之间, 同时再乘以姿势矩阵 M_j 。因此, 对熵值 S_j 做一些简单的尺度变换, 然后使用 Sigmoid 函数激活, 得到激活值 a_j :

$$a_j = \text{Sigmoid}(\lambda(\beta_\alpha - (\beta_\alpha + \sum_{j=1}^d \ln(W_{ij} M_j)) \sum_i r_{ij})) \quad (12)$$

$$\lambda = \lambda_{\text{final}} * (1 - 0.95^{r+1}) \quad (13)$$

其中, r 表示迭代次数, λ_{final} 设置为 0.01。

综上, 下面给出使用 SFM 聚类算法实现胶囊网络中的路由迭代过程的算法流程, 如算法 1 所示。

算法 1 胶囊网络中 SFM 算法的迭代过程

输入: 精度 e_1 和 e_2 , 指数 m , 迭代次数 r

输出: 隶属度矩阵 W_{ij} , 激活值 a_j

Step1 对 $\forall i \in \Omega_L, j \in \Omega_{L+1}: W_{ij} \leftarrow 1 / |\Omega_{L+1}|$;

Step2 平均初始化隶属度矩阵 W_{ij} , 随机初始化激活值 $a_j (0 < a_j \leq 1)$;

Step3 迭代 r 次:

Step4 $v_j \leftarrow \sum_{i=1}^n a_j W_{ij}^m u_i / \sum_{i=1}^n a_j W_{ij}^m$

Step5 $W_{ij} \leftarrow 1 / \sum_{c=1}^k (\|u_i - v_c\| / \|u_i - v_c\|)^{(2/m-1)}$

Step6 $R_{ij} \leftarrow a_j W_{ij} / \sum_{j=1}^k a_j W_{ij}$

Step7 $r_{ij} \leftarrow R_{ij} / \sum_{i=1}^n R_{ij}$

Step8 $\text{cost}_j \leftarrow (\beta_\mu + \sum_{l=1}^d \ln(W_{ij} M_j)) \sum_i a_i^{\text{last}} r_{ij}$

Step9 $a_j \leftarrow \text{Sigmoid}(\lambda(\beta_\alpha - \text{cost}_j))$

Step10 $J' \leftarrow \sum_{j=1}^k \sum_{i=1}^n W_{ij}^m \|u_i - v_j\|^2$

Step11 如果 $J' - J'^{-1} < e_1$

Step12 退出一次迭代

Step13 $m \leftarrow m + 1$

Step14 继续一轮上述迭代过程

Step15 如果 $J'_T - J'_{T-1} > e_2$ 或迭代次数大于 r

Step16 返回上一轮 W_{ij} 和 a_j 的值

其中, J' 为此次迭代的目标函数的值, J'_T 为此轮目标函数的最终值。

3.2 多尺度特征融合独立训练

胶囊网络中的第一步就是利用卷积层进行特征提取的过程, 然后通过胶囊层将浅层特征逐步整合成更深层的特征并最终用于分类器中。通过卷积神经网络提取的深层特征往往具有较强的语义特征, 但分辨率较低; 而浅层特征虽然包含的语义信息较少, 但具有较强的局部细节, 且保留了一定的空间信息。传统的神经网络一般将最后一个卷积层提取出的特征作为分类器的输入, 但越来越多的研究表明, 保留了空间信息的浅层特征同样对深度学习中的模型识别任务有一定帮助, 将不同尺度的特征融合在一起, 可以使得融合后的特征包含更多有用的信息。特征金字塔网络以自底向上的方式使用横向连接将不同尺度的特征融合在一起并进行独立训练, 这种训练方法在没有加剧计算损耗的同时充分利用了多尺度特征的特点, 增强了特征的表达能力, 提升了模型的效果。

借鉴特征金字塔网络多尺度特征融合独立训练的思想, 对于具有 n 个胶囊层 (n 次路由迭代) 的胶囊网络, 我们保留经过每一个胶囊层后的特征。最终, 将胶囊层提取出来的特征首先输送到全连接层 (使用 Softmax 分类器) 做一次预测, 然后将其上采样成与 $n-1$ 胶囊层特征同样的尺度, 并以 element-wise 的方式进行点积相乘来融合它们。该 $n-1$ 胶囊层的特征图需要先进行一个 1×1 的卷积以降维, 然后将融合后的特征输送到全连接层再做一次预测。如果以 $n-2$ 胶囊层的特征为标准, 则需要融合最后 3 个胶囊层的特征, 即具有 n

个胶囊层的胶囊网络最多允许做 n 次特征融合独立训练。一个使用最后 3 个胶囊层的多尺度特征融合独立训练的过程如图 3 所示。

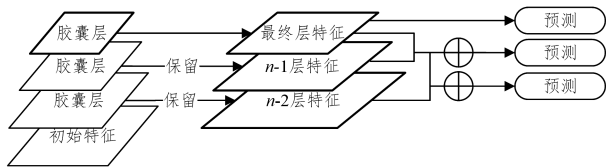


图 3 使用最后 3 个胶囊层的特征融合独立训练过程

Fig. 3 Fusion independent training using features of last three capsule layers

使用最后 3 层的特征进行融合独立训练的过程可以用如下公式来表示:

$$y_1 = \text{Softmax}(Cap^n) \quad (14)$$

$$y_2 = \text{Softmax}(\text{Unsamp}(Cap^n) \oplus Cap^{n-1}) \quad (15)$$

$$y_3 = \text{Softmax}(\text{Unsamp}(Cap^n) \oplus \text{Unsamp}(Cap^{n-1}) \oplus Cap^{n-2}) \quad (16)$$

式(14)表示使用最终胶囊层(第 n 层)的特征进行训练,式(15)表示使用最后两层的特征进行融合训练,式(16)则表示使用最后 3 层的胶囊层特征进行融合训练。其中, Cap^n 表示最终胶囊层的输出特征, Cap^{n-1} 表示上一胶囊层的输出特征, $\text{Unsamp}(\ast)$ 函数表示上采样操作, \oplus 表示相同大小的特征以 element-wise 的方式进行点积相乘的融合过程, $\text{Softmax}(\ast)$ 函数表示输送到全连接层的特征使用 Softmax 分类器进行分类。这个融合独立训练的过程是弹性的,即根据输入样本的不同,可以有选择地融合特定胶囊层的输出特征。

由于第一次训练能够获得一个较粗略的模型,因此在后面以低层胶囊特征为标准进行融合独立训练的过程中,将学习率降低一个量级以微调模型的参数。

4 实验及结果分析

为了证明将根据 SFCM 算法实现的路由迭代过程和多尺度特征融合独立训练应用到胶囊网络中的有效性,本节基于 Keras 框架在 MNIST 和 CIFAR-10 数据集上验证本文的模型。

4.1 数据集介绍

MNIST 是一个具有 10 个数字的手写体数字识别数据集,每张图片为 28×28 像素的灰度图像,共包含 60 000 张训练集图像和 60 000 个训练集标签、10 000 张测试集图像和 10 000 个测试集标签,其中 60 000 张训练集图像被拆分为包含 55 000 个图像的训练集和包含 5 000 个图像的验证集,所有图像的标签均被初始化为 10 维的“one-hot”向量。MNIST 数据集是图像分类领域的一个简单而经典的数据集,由于在其上的训练速度比较快,常被用于粗略评估模型的性能。

CIFAR-10 是一个具有多种背景信息的带标签数据集,与 MNIST 一样具有 10 种类别的物体,由 32×32 像素的 RGB 图像组成。每个类有 6 000 张图像,总共分为 50 000 张训练集图像和 10 000 张测试集图像,测试集包含从每个类别中随机选择的 1 000 张图像,采用交叉验证的方式。

4.2 模型细节

实验基于 Keras 框架设计。在 MNIST 数据集上,为了便于将卷积神经网络模型与胶囊网络模型进行对比,我们将原本的 10 分类问题转换成 10 个二分类问题。两种模型的 CNN 部分保持一致,先经过 2 个 64 维的 3×3 卷积,然后进行 2×2 的平均池化,再经过 2 个 128 维的 3×3 卷积,上述过程中的激活函数均为 ReLU。在此之后,卷积神经网络模型连接一个 128 维的使用 ReLU 激活函数的全连接层和一个 10 维的使用 Sigmoid 激活函数的全连接层。胶囊网络模型则将池化之后的部分替换为相应的胶囊结构,最终胶囊层为 10 个 16 维的胶囊单元,进行 3 次路由迭代,并保留每次路由迭代后整合的特征,最多允许 20 次 epoch,使用 Keras 默认的 Adam 优化器,将测试准确率作为 MNIST 数据集上的评价指标。

在 CIFAR-10 数据集的实验中,将图像缩放到 28×28 像素,使得在 CIFAR-10 数据集上的模型与在 MNIST 数据集上的单个模型大致具有相同的架构,主要的区别是前者增加了 3 个颜色通道。为了应对 CIFAR-10 数据集中更加复杂的图像特征,将 3 次路由迭代中胶囊层单元的个数分别修改为 8, 12, 16, 这种增加高级胶囊层维度的方式能够应对更加复杂的融合特征。我们还进行了 4 次路由迭代的实验(最终胶囊层中单元的个数增加到 24),并将测试错误率作为 CIFAR-10 数据集上的评价指标。模型的损失函数为:

$$\text{Loss} = y * \max(0.9 - \hat{y}, 0)^2 + 0.25(1 - y) * \max(\hat{y} - 0.1, 0)^2 \quad (17)$$

其中, y 为样本的真实值, \hat{y} 为预测值。将正例预测与负例预测的比例设置为 4:1,使得在以正例预测为主导对模型进行调整时,也能够充分利用负例预测对模型的调整作用。

4.3 结果分析

4.3.1 MNIST 上的实验结果

为了充分证明胶囊网络本身的泛化能力,训练完模型后,在测试阶段随机挑选两张测试集的图片进行拼接,我们把该测试集命名为 Jnt-MNIST,并且分别对比了不考虑置信度(取得分最高的两个类别)和置信度为 0.5(取得分最高的两个类别且这两个最高得分的类别分数都需要超过 0.5)的两种情况。结果发现,传统的卷积神经网络模型在对拼接图像进行测试时效果骤降,在不考虑置信度时准确率为 42.31%,考虑置信度时准确率则只有 10.57%;而文献[14]中的胶囊网络模型在两种情况下的准确率均超过了 90%,实验结果如表 2 所列。

表 2 MNIST 上的实验结果

Table 2 Experimental results on MNIST

	MNIST	Jnt-MNIST	Jnt-MNIST(置信度 0.5)
CNN	99.32	42.31	10.57
Capsule	99.37	93.72	91.90
FM-Capsule	99.26	91.88	91.11
F-Capsule	99.37	94.13	92.26

这两种模型均没有使用数据增强等技巧,在没有进行针对性训练的情况下, Jnt-MNIST 上的实验结果证明了胶囊网

络具有比传统的卷积神经网络更强的泛化能力。这种天生能够将学习到的知识进行迁移拓展的能力,正是一直以来卷积神经网络发展历程中致力于解决的问题。

本文提出的使用模糊聚类的多尺度特征融合独立训练的胶囊网络(FM-Capsule)也在这3种形态的MNIST数据集上进行了实验,训练过程中使用变化的学习率,在以更低胶囊层特征为标准进行融合训练时,将学习率降低一个量级以微调模型参数。在测试阶段就发现FM-Capsule(使用3层特征融合)的准确率均低于文献[14]中模型的结果。考虑到原生的MNIST是一种构成简单且没有背景干扰的数据集,使用多尺度特征融合独立训练可能会给原本就比较简单清晰的特征图增添更多复杂的信息而降低准确率,我们去掉了多尺度特征融合训练部分,使用仅更改聚类算法为SFCM算法的胶囊网络(F-Capsule)进行验证。在不考虑置信度的Jnt-MNIST上,F-Capsule模型获得了94.13%的准确率,在置信度为0.5的Jnt-MNIST上获得了92.26%的准确率,准确率均超过了使用K-Means算法作为特征整合方案的文献[14]中的结果,这证明了模糊聚类算法作为一种普适性更强的聚类算法应用到胶囊网络中的潜力。

接下来,为了进一步证明SFCM算法应用在胶囊网络中的效果,并检验多尺度特征融合独立训练是否对胶囊网络具有提升作用,我们在更加复杂的图像数据集CIFAR-10上进行了实验。

4.3.2 CIFAR-10上的实验结果

本节的主要目的是验证将多尺度特征融合独立训练应用在胶囊网络上对模型产生的影响,这其中涉及到动态次数选取的问题。文献[14]在CIFAR-10数据集上对路由迭代次数进行了探索,实验结果表明当进行5次路由迭代时,模型训练损失的变化几乎可以忽略不计。考虑到网络的容量和多次的路由迭代会增加模型过度匹配训练集的风险,文献[14]建议在胶囊网络中均使用3次路由迭代。而由于本文模型进行特征提取的卷积层比文献[14]更轻量,初次提取出的特征相对简单,为了加强特征表示,并增加多尺度特征融合独立训练方法中方案的选取数量,本文将动态路由增加为4次。我们进行了融合不同的胶囊层特征的对比实验,也将本文的胶囊网络模型与文献[14]和文献[15]中的模型进行了对比,实验结果如表3所列。

表3 CIFAR-10上的实验结果

Table 3 Experimental results on CIFAR-10

模型	聚类方式	多尺度	路由次数	融合层数	测试错误率/%
CapsNet (K-means)	K-means	否	3		11.88
CapsNet (GMM)	GMM	否	3		11.90
F-Capsule	FCM	否	3		11.62
FM-Capsule	FCM	是	3	1,2,3	11.59
FM-Capsule	FCM	是	3	2,3	11.59
FM-Capsule	FCM	是	3	1,3	11.61
FM-Capsule	FCM	是	4	1,2,3,4	11.65
FM-Capsule	FCM	是	4	2,3,4	11.48
FM-Capsule	FCM	是	4	1,3,4	11.50
FM-Capsule	FCM	是	4	3,4	11.54
FM-Capsule	FCM	是	4	2,4	11.48
FM-Capsule	FCM	是	4	1,4	11.51

最好的实验结果出现在动态路由次数为4次时,在融合第2,3,4层特征和融合第2,4层特征的情况下,5次结果的平均测试错误率达到了11.48%。我们分析出现这种现象出现的原因可能是:第2层特征包含的空间信息较第3层更为清晰,再同与包含了高级语义特征的第4层特征进行融合,能够获得综合了双方优点的融合特征,而此时第3层特征则具有很小的影响。我们发现将最终层特征与第1层特征或第2层特征进行融合时的实验结果相对接近,这表明最终层特征对模型的表现起到了最主要的作用。另外,我们发现融合了全部4层特征时的测试错误率反而升高到了11.65%,经分析,我们认为这种情况可能是过拟合导致的。因此,我们比较了动态路由次数为4时,分别使用最后3层特征融合和全部4层特征融合时模型的训练损失,如图4所示。实验结果表明,尽管使用4层特征融合的训练损失在每个阶段均低于使用3层特征融合的训练损失,但在测试阶段,使用最后3层特征融合的网络模型在CIFAR-10上的测试错误率反而小于使用4层特征融合的网络模型。以上说明当迭代次数较大时,使用过多的层数进行特征融合训练可能会导致最终的网络模型过于拟合训练集,从而影响了其在测试集上的表现,这表明有策略地选择路由迭代的次数和需要进行特征融合的层数是必要的。

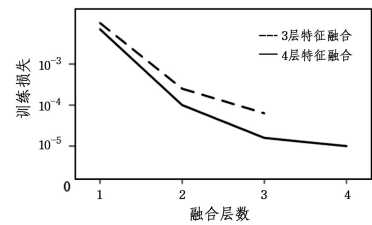


图4 3层特征融合和4层特征融合的训练损失

Fig. 4 Training loss of three-layer feature fusion and four-layer feature fusion

实验结果还表明,单引入SFCM算法,F-Capsule在CIFAR-10上的效果就超过了文献[14]和文献[15]中的胶囊网络模型的效果,从而证明了在胶囊网络中处理具有复杂背景的图片时,本文提出的SFCM算法是一种比原先的聚类算法更加合适的特征整合方案,也体现了模糊聚类算法的普适性。文献[14]中,10.60%的测试错误率是建立在7个模型的集合上实现的,我们采用文献[15]中使用的K-means算法作为特征整合方案,在单个模型上实现了11.88%的测试错误率,该结果低于使用SFCM算法作为特征整合方案的F-Capsule模型上达到的11.62%。

我们还发现,引入了多尺度特征融合独立训练方法之后,使用3次路由迭代的FM-Capsule相比于F-Capsule的提升很小,这可能是由于本文模型在特征提取阶段使用的卷积网络层数较少且通道数量不如文献[14]中的多,初始提取的特征比较粗糙,但更少的参数也加快了模型的训练速度。考虑到计算成本和时间因素,我们将重点放在胶囊网络结构变化的对比之上,因此并未花费过多时间去调整提取初始特征的卷积层。

结束语 胶囊网络是一种新型网络结构,它将特征表示由标量提升为向量,这种表示方式极大地提升了特征的表

达能力,使得胶囊网络具有很好的泛化能力,在 Jnt-MNIST 数据集上对胶囊网络模型和卷积神经网络模型进行对比,实验结果证明了这一点;并且相较于传统的卷积神经网络,胶囊网络使用较少的胶囊层和相对更少的参数量就能够达到标准卷积神经网络在图像分类上的效果。由于胶囊网络总是会考虑图像中的所有信息,无法很好地应对包含了复杂背景信息的数据集,因此其在图像任务上的表现不如一些最新的深度学习方法。但是,作为一种新型的网络,它在 CIFAR-10 上的测试错误率也超过了标准卷积神经网络首次应用到 CIFAR-10 上的结果,且超过了绝大多数参数量相近或训练时间相当的卷积神经网络模型,这些都说明胶囊网络具有很大的探索空间,提高其对复杂背景的鲁棒性是将来的主要工作之一。另外,无论在胶囊网络中使用 K-Means 聚类算法或 GMM,还是本文提出的 SFCM 算法,每个胶囊层选取的单元个数(即聚类中心的个数)都是一个超参数,而一些决定聚类中心个数的策略又不太适合应用到胶囊网络中,由于在图像特征比较复杂时通过那些策略选取的聚类中心个数往往会很大,这会极大地加剧网络的负担,因此如何更好地决定每一胶囊层中单元的个数也是一个值得研究的课题。

参 考 文 献

- [1] LU D, WENG Q. A Survey of Image Classification Methods and Techniques for Improving Classification Performance[J]. *International Journal of Remote Sensing*, 2007, 28(5): 823-870.
- [2] YANN L, BOTTOU L, BENGIO Y, et al. Gradient-based Learning Applied to Document Recognition[J]. *Proceedings of the IEEE*, 1998, 86(11): 2278-2324.
- [3] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet Classification with Deep Convolutional Neural Networks[C]// NIPS. New York, NY: Curran Associates, 2012: 1097-1105.
- [4] ZEILER M D, FERGUS R. Visualizing and Understanding Convolutional Networks[C]// European Conference on Computer Vision. Switzerland: Springer, 2014: 818-833.
- [5] SIMONYAN K, ZISSERMAN A. Very Deep Convolutional Networks for Large-scale Image Recognition[C]// Computer Vision and Pattern Recognition. Princeton: CVPR Press, 2015.
- [6] SZEGEDY C, LIU W, JIA Y Q, et al. Going Deeper with Convolutions[C]// CVPR. Piscataway, NJ: IEEE, 2015: 1-9.
- [7] HE K, ZHANG X Y, REN S Q, et al. Deep Residual Learning for Image Recognition [C] // CVPR. Piscataway, NJ: IEEE, 2016: 770-778.
- [8] HUANG G, LIU Z, MAATEN L, et al. Densely Connected Convolutional Networks[C]// Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2017: 2261-2269.
- [9] GU Z B, CAO F L. Image Classification algorithm of Multi-layer Forward Artificial Neural Network [J]. *Computer Science*, 2018, 45(S2): 238-243. (in Chinese)
顾哲彬, 曹飞龙. 多层前向人工神经网络图像分类算法[J]. *计算机科学*, 2018, 45(S2): 238-243.
- [10] HINTON G E, SRIVASTAVA N, KRIZHEVSKY A, et al. Improving Neural Networks by Preventing Co-adaptation of Feature Detectors[C/OL]. <http://arxiv.org/abs/1207.0580>.
- [11] IOFFE S, SZEGEDY C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift[C]// International Conference on International Conference on Machine Learning. New York, NY: ACM, 2015: 448-456.
- [12] TONG W G, LI M X, ZHANG Y K. Research on Optimization Algorithm of Deep Learning [J]. *Computer Science*, 2018, 45(S2): 155-159. (in Chinese)
全卫国, 李敏霞, 张一可. 深度学习优化算法研究[J]. *计算机科学*, 2018, 45(S2): 155-159.
- [13] FU B W, SUN T, LIANG J, et al. Review of Principle and Application of Deep Learning[J]. *Computer Science*, 2018, 45(S1): 11-15. (in Chinese)
付文博, 孙涛, 梁藉, 等. 深度学习原理及应用综述[J]. *计算机科学*, 2018, 45(S1): 11-15.
- [14] SABOUR S, FROSST N, HINTON G E. Dynamic Routing between Capsules[C]// NIPS. New York, NY: Curran Associates, 2017: 3859-3869.
- [15] HINTON G E, SABOUR S, FROSST N. Matrix Capsules with EM Routing[C]// International Conference on Learning Representations. Princeton: ICLR Press, 2018.
- [16] JAISWAL A, ABDALMAGEED W, WU Y, et al. CapsuleGAN: Generative Adversarial Capsule Network[C]// European Conference on Computer Vision. Berlin: Springer, 2018: 526-535.
- [17] DENG F, PU S L, CHEN X H, et al. Hyperspectral Image Classification with Capsule Network Using Limited Training Samples[J]. *Sensors*, 2018, 18(9): 3153.
- [18] ZHANG X N, LUO P C, HU X W, et al. Research on Classification Performance of Small-Scale Dataset Based on Capsule Network[C]// International Conference on Robotics and Artificial Intelligence. New York, NY: ACM, 2018: 24-28.
- [19] XIANG C Q, ZHANG L, TANG Y, et al. MS-CapsNet: A Novel Multi-Scale Capsule Network[J]. *IEEE Signal Processing Letters*, 2018, 25(12): 1850-1854.
- [20] BEZDEK J C, EHRlich R, FULL W. FCM: The Fuzzy c-means Clustering Algorithm[J]. *Computers & Geosciences*, 1984, 10(2/3): 191-203.
- [21] ADELSON E H, ANDERSON C H, BERGEN J R, et al. Pyramid Methods in Image Processing [J]. *RCA engineer*, 1984, 29(6): 33-41.
- [22] BELL S, ZITNICK C L, BALA K, et al. Inside-Outside Net: Detecting Objects in Context with Skip Pooling and Recurrent Neural Networks[C]// International Conference on Learning Representations. Piscataway, NJ: IEEE, 2016: 2874-2883.
- [23] KONG T, YAO A B, CHEN Y R, et al. HyperNet: Towards Accurate Region Proposal Generation and Joint Object Detection [C]// International Conference on Learning Representations. Piscataway, NJ: IEEE, 2016: 845-853.
- [24] LIN T Y, DOLLAR P, GRISHICK R, et al. Feature Pyramid Networks for Object Detection [C]// International Conference on Learning Representations. Piscataway, NJ: IEEE, 2017: 936-944.
- [25] FRANCIS D, HUET B, MERALDO B. Embedding Images and Sentences in a Common Space with a Recurrent Capsule Network[C]// International Workshop on Content-based Multimedia Indexing. Piscataway, NJ: IEEE, 2018: 1-6.