

# 不平衡数据对卷积神经网络的影响及改进算法

马 骏,钱亚冠,郭艳凯,吴淑慧,云本胜

(浙江科技学院 曙光大数据学院,杭州 310023)

**摘 要:** 通过不同的数据分布、激活函数和网络结构对卷积神经网络(convolutional neural networks,CNN)的训练过程进行试验分析发现,数据不平衡会造成CNN训练过程收敛慢、泛化能力差的负面影响。针对这一问题,结合过抽样和欠抽样各自的优点,在随机梯度下降算法的基础上,提出均衡小批量随机梯度下降算法(equilibrium mini-batch stochastic gradient descent,EMSGD),保证小批量内的数据均衡,精确调整更新参数的梯度方向。试验结果表明,均衡小批量随机梯度下降算法可以在数据不平衡条件下提高CNN训练误差收敛速度,提高泛化性能。

**关键词:** 数据不平衡;卷积神经网络;随机梯度下降

**中图分类号:** TP183      **文献标志码:** A      **文章编号:** 1671-8798(2020)03-0181-10

## Influence upon and improved algorithm of convolutional neural networks under data imbalance

MA Jun, QIAN Yaguan, GUO Yankai, WU Shuhui, YUN Bensheng

(School of Sugon Big Date Science, Zhejiang University of Science and Technology, Hangzhou 310023, Zhejiang, China)

**Abstract:** Experiments under different data, activation functions and network structures show that the CNN training error converges slowly and the generalization ability is poor under imbalanced training data. In response to this problem, in combination with advantages of over-sampling and under-sampling, equilibrium mini-batch stochastic gradient descent(EMSGD) was put forward on the basis of the mini-batch stochastic gradient descent, ensuring the data balance in mini-batch and adjusting accurately the gradient direction of update parameters. Experiments prove that EMSGD can raise convergence speed of CNN training error under the condition of imbalanced training data and improve the generalization ability.

**Keywords:** data imbalance; convolutional neural networks(CNN); stochastic gradient descent

---

**收稿日期:** 2019-05-21

**基金项目:** 浙江省公益技术应用研究项目(LGG19F030001);浙江省自然科学基金项目(LY17F020011);国家自然科学基金项目(61572163)

**通信作者:** 钱亚冠(1976—),男,浙江省嵊州人,副教授,博士,主要从事机器学习和神经网络研究。E-mail: qianyaguan@zust.edu.cn。

随着深度学习的兴起和人工智能第三次浪潮的到来,深度神经网络被成功地应用到图像识别<sup>[1-2]</sup>、语音识别<sup>[3-4]</sup>、自然语言处理等领域<sup>[5-6]</sup>。在 2015 年 ImageNet 项目举办的 ImageNet 大规模视觉识别挑战赛(ImageNet large scale visual recognition challenge, ILSVRC)中,微软研究院团队提出的深度学习模型成功地将图像识别错误率降低到约 3.57%,低于人类 5.1%的错误率<sup>[1]</sup>。微软研究院团队的成功极大地激发了科研人员对深度神经网络的研究热情。但在研究过程中我们发现,训练数据不平衡会导致深度卷积神经网络(convolutional neural networks, CNN)难以收敛及泛化能力差等问题。

数据不平衡现象广泛存在于生物医疗<sup>[7]</sup>、金融<sup>[8]</sup>、信息安全<sup>[9]</sup>、工业<sup>[10]</sup>、计算机视觉<sup>[11]</sup>等诸多领域。机器学习领域对数据不平衡问题已经做了大量研究<sup>[12-13]</sup>,目前主要有数据层面和算法层面两种解决方法<sup>[14]</sup>。数据层面主要通过改变训练集的类分布来实现数据均衡,典型的方法有过抽样<sup>[15]</sup>和欠抽样<sup>[16]</sup>。过抽样方法首先复制随机抽取的少数类样本,并将复制的样本加入原本的数据集,从而使原本的数据集各类别数据平衡;与过抽样相反,欠抽样方法将多数类数据从原始数据集中抽离来使原始数据集各类比数据平衡。有研究表明:过抽样方法比较稳定有效,但容易引起过拟合<sup>[15]</sup>;欠抽样方法的缺点是丢失了多数类的很多重要信息,但在有的情况下欠抽样优于过抽样<sup>[16]</sup>。算法层面的方法典型的有代价敏感学习<sup>[17]</sup>,通过给多数类和少数类赋予不同的误分类代价,从而改善数据不平衡问题。

数据不平衡不仅对传统分类器的训练造成负面影响,而且对多层感知器同样会影响训练阶段的收敛和测试阶段的泛化<sup>[18]</sup>。在深度学习中最常用的解决方法仍然是过抽样技术<sup>[19-20]</sup>,此外代价敏感学习也被应用到解决深度神经网络上的数据不平衡问题<sup>[21-22]</sup>。最新研究提出的神经网络训练方法将网络训练分为 2 个阶段,先用均衡数据进行训练,再在不均衡数据上对输出层进行微调<sup>[23]</sup>。总体而言,目前研究人员对深度学习的不均衡问题及其处理方法还没有较为系统的分析,只是使用了一些基于直觉的解决方法和解决传统机器学习不平衡问题的技术。

基于上述研究,本文主要研究不平衡训练数据对 CNN 的影响,通过试验发现 CNN 在不同激活函数和不同网络结构情况下用多种类型的不均衡数据进行训练时,均会造成模型训练过程收敛缓慢、模型准确率低的现。因此,我们提出均衡小批量随机梯度下降算法(equilibrium mini-batch stochastic gradient descent, EMSGD),在每一轮梯度迭代过程中,从每类数据中抽取相同数量的样本组成小批量,从而保证小批量内的数据均衡。试验结果表明,我们提出的方法在 CNN 中可实现快速收敛并获得更好的泛化能力。

## 1 理论基础

CNN 是一类包含卷积计算且具有深度结构的前馈神经网络,在诸多领域特别是在图像相关领域起着重要作用。CNN 有多种常用的模型,LeNet-5<sup>[24]</sup>是最早的 CNN 模型之一,其结构简单,只有 2 个卷积层、2 个池化层和 2 个全连接层。虽然其层数较少,但包含了 CNN 的基本结构,即卷积-池化-全连接的网络结构,为深度卷积神经网络的发展奠定了良好的基础。之后的 AlexNet 等<sup>[2]</sup>CNN 模型均建立在 LeNet-5 基础上,同样具有卷积-池化-全连接的网络结构。AlexNet 的网络结构如图 1 所示。

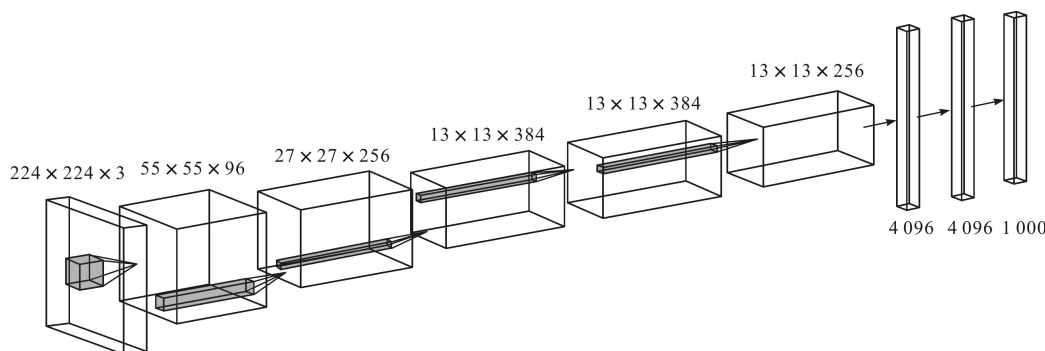


图 1 AlexNet 网络结构

Fig. 1 AlexNet network architecture

## 1.1 CNN 各层作用

卷积层是 CNN 的核心,其实质是两个矩阵的计算,即代表图像的矩阵和代表卷积核的矩阵之间对应位置相乘再求和。通过激活函数对结果进行非线性变换,以提取图像特征,得到特征图。卷积层通过权重共享减少 CNN 参数量,便于训练。

池化层又称采样层,一般在卷积层后面,常见的池化运算有最大池化和平均池化。池化操作是将卷积得到的特征图与池化核覆盖区域取平均值或取最大值作为池化结果,具有空间不变性和特征降维的作用,在一定程度上可防止过拟合。卷积和池化一般具有多层,即卷积后进行池化,池化后再卷积,周而复始,以提取原始图像的复杂特征。

多层卷积池化后是全连接层,其作用是将卷积池化层提取到的特征映射到样本的标记空间,对输入样本进行分类。对于多分类问题,最后一层的全连接层一般为 Softmax 层,即对输出进行 Softmax 计算,用以输出概率向量,并且计算梯度时更为简单。

## 1.2 激活函数

选择不同的激活函数会影响 CNN 性能和训练过程,常用的激活函数有 Sigmoid( $S$ )函数、Tanh( $T$ )函数和 Relu( $R$ )函数等,定义如下:

$$S(z) = \frac{1}{1 + e^{-z}} \quad (1)$$

$$T(z) = \frac{1 - e^{-2z}}{1 + e^{-2z}} \quad (2)$$

$$R(z) = \begin{cases} 0, & z < 0; \\ z, & z \geq 0. \end{cases} \quad (3)$$

式(1)~(3)中: $z$ 为激活函数的输入。 $S$ 函数和 $T$ 函数相似,均呈指数形式,但 $S$ 函数值域为 $(0,1)$ , $T$ 函数为 $(-1,1)$ 。因为 $T$ 函数的值关于原点对称,而且幅度要比 $S$ 函数大,一阶导数也比 $S$ 函数的更大,所以使用 $T$ 函数的网络在训练时,收敛速度更快。然而 $S$ 函数与 $T$ 函数均存在梯度饱和区,即函数值接近边界时梯度接近0,训练时可能会出现梯度消失问题,导致训练困难。随着神经网络层数的加深,梯度消失问题更容易出现,所以可以避免梯度消失的 $R$ 函数被更多地应用在具有多层结构的深度神经网络中,例如 ResNet<sup>[1]</sup>、AlexNet<sup>[2]</sup>等,并且其稀疏性特点可以缓解模型过于复杂时可能会出现过的拟合问题。

## 2 数据不平衡对 CNN 的影响

为探讨数据不平衡对 CNN 的影响,抽取 MNIST (Mixed National Institute of Standards and Technology)数据集部分数据构建多种不平衡的数据以训练 CNN 模型,并采用不同的激活函数。试验结果表明,在不同激活函数和网络结构情况下,不平衡训练数据均会对 CNN 造成不良影响,如对少数类的测试准确率偏低,其总体准确率也低于均衡数据情况下得到的 CNN。

### 2.1 数据设置

MNIST 是一个被广泛应用于机器学习性能测试的手写体数字数据集,共包含 10 个类别,为 0~9,图像均为灰度图像,由  $28 \times 28$  个像素点构成,图 2 为 MNIST 数据集手写体数字实例。从中抽取数据构建 4 种各类数据比例不同的训练数据分布,各训练数据分布总数均为 10 000 张图片,其中分布 1 为数据均衡的情况,各类数据均为 1 000 张图片;分布 2 为一类数据多于另一类数据;分布 3 为某一类数据多于其余类数据;分类 4 为各类数据量呈线性增长;测试数据由 MNIST 划分,总数为 10 000 张图片,各类数据数量相近。MNIST 数据集不同分布的训练数据见表 1。



图 2 MNIST 数据集手写体数字实例  
Fig. 2 Examples of handwritten numerals in MNIST datasets

表 1 MINIST 数据集不同分布的训练数据

Table 1 MINIST training data with different distribution

张

分布	各类图像数量									
	0	1	2	3	4	5	6	7	8	9
1	1 000	1 000	1 000	1 000	1 000	1 000	1 000	1 000	1 000	1 000
2	1 300	700	1 300	700	1 300	700	1 300	700	1 300	700
3	1 900	900	900	900	900	900	900	900	900	900
4	550	650	750	850	950	1 050	1 150	1 250	1 350	1 450
测试数据	980	1 135	1 032	1 010	982	892	958	1 028	974	1 009

## 2.2 网络和训练参数设置

在数据集较为简单、数据量较少的情况下,模型过于复杂易导致过拟合,因此我们采用具有完整 CNN 结构但较为简单的一层卷积(卷积核大小为  $5 \times 5$ ,数量为 20)、一层池化(最大池化,池化核大小为  $2 \times 2$ )和两层全连接(神经元数分别为 100、10)的 CNN,最后一层全连接用 Softmax 输出结果。使用小批量随机梯度下降(mini-batch stochastic gradient descent,MSGD)算法进行训练,学习率为 0.01,小批量大小为 200,迭代次数为 20 000,代价函数使用交叉熵代价函数,激活函数使用 S 函数、T 函数和 R 函数分别进行试验。为了对比不同网络结构下的试验结果,在具有 2 个卷积池化层和 2 个全连接层的经典 CNN 结构 LeNet-5 上进行重复试验,激活函数使用 R 函数,代价函数为交叉熵代价函数,对网络参数做了一些调整,其余训练参数与上述的相同。本试验的硬件条件为:Win10 64 位系统,3.30 GHz CPU,4 G 内存等。在 Python 平台进行试验,并借助 Theano 库完成高性能计算。

## 2.3 试验验证

在不同训练数据和不同激活函数情况下由 MSGD 训练所得模型的 CNN 网络分类准确率见表 2~4。表 5 是 LeNet-5 模型在不同训练数据下由 MSGD 训练所得模型的分分类准确率。

观察在 S 函数下各类别的分类准确率,发现大部分情况下训练数据越多的类别在测试阶段分类准确率越高,如表 2 分布 2 中偶数类,分布 3 中 0 类,分布 4 中 8、9 类。由表 2 可知,均衡数据(分布 1)情况下训练得到的 CNN 模型分类准确率明显优于不均衡数据(分布 2~4)情况下的分类准确率。而在相同的数据和网络结构情况下,采用不同的激活函数,发现也存在类似的情况。由表 3 和表 4 可知,在相同激活函数下,均衡数据(分布 1)训练得到的 CNN 模型测试准确率要高于不均衡数据(分布 2~4)。由表 5 可知,不均衡训练数据同样会对不同的 CNN 模型造成影响。虽然 LeNet-5 模型在不同数据分布情况下得到的模型分类准确率较高,可以在一定程度上降低不均衡数据带来的不良影响,但是均衡数据(分布 1)情况下得到的 CNN 模型分类准确率还是明显优于不均衡数据(分布 2~4)情况下的分类准确率。

表 2 由 MSGD 训练的激活函数为 S 函数的 CNN 网络分类准确率

Table 2 Classification accuracy of CNN networks as S activation function trained by MSGD

%

分布	测试数据 准确率	手写体数字数据准确率										训练数据 准确率
		0	1	2	3	4	5	6	7	8	9	
1	88.00	98.00	88.67	86.22	77.56	88.89	90.83	90.00	78.00	92.56	90.44	88.72
2	69.94	98.11	67.67	93.56	41.00	95.78	40.33	93.56	68.22	91.67	13.33	77.19
3	81.14	99.78	85.00	79.78	59.56	83.56	78.67	84.11	85.44	73.11	81.67	83.89
4	71.55	86.33	72.89	53.44	47.44	36.67	60.33	92.44	71.44	95.44	96.44	75.26

表 3 由 MSGD 训练的激活函数为 T 函数的 CNN 网络分类准确率

Table 3 Classification accuracy of CNN networks as T activation function trained by MSGD

%

分布	测试数据 准确率	手写体数字数据准确率										训练数据 准确率
		0	1	2	3	4	5	6	7	8	9	
1	93.97	98.56	96.00	93.67	88.44	93.78	94.33	95.00	89.11	96.44	93.33	96.73
2	90.65	99.00	91.56	97.11	79.89	97.89	86.33	96.78	85.22	96.44	74.33	94.46
3	93.85	99.56	85.00	79.78	59.56	83.56	78.67	84.11	85.44	73.11	81.67	96.77
4	90.73	95.44	92.11	88.11	81.78	78.78	91.17	94.78	90.78	97.56	97.00	93.67

表 4 由 MSGD 训练的激活函数为  $R$  函数的 CNN 网络分类准确率Table 4 Classification accuracy of CNN networks as  $R$  activation function trained by MSGD %

分布	测试数据 准确率	手写体数字数据准确率										训练数据 准确率
		0	1	2	3	4	5	6	7	8	9	
1	95.38	98.44	97.67	95.22	93.33	95.22	95.17	96.33	89.89	97.78	93.89	97.78
2	90.79	99.22	92.22	98.33	73.22	97.89	89.33	97.22	86.56	97.22	76.22	94.52
3	94.78	99.78	98.00	94.44	90.89	95.22	93.67	95.44	90.56	95.44	92.67	97.51
4	88.98	88.11	90.89	81.89	81.44	77.44	89.00	97.11	88.11	96.67	98.11	91.63

表 5 LeNet-5 网络 MSGD 训练的分类准确率

Table 5 Classification accuracy of MSGD training for LeNet-5 networks %

分布	测试数据 准确率	手写体数字数据准确率										训练数据 准确率
		0	1	2	3	4	5	6	7	8	9	
1	96.18	98.44	97.67	98.00	96.67	96.44	97.83	95.56	87.78	96.11	95.89	98.74
2	93.29	99.22	96.44	98.78	84.22	98.11	92.00	98.22	85.44	96.89	82.44	96.82
3	94.37	99.78	95.33	97.89	91.44	95.56	95.83	96.11	86.33	96.78	88.78	96.67
4	91.77	93.78	95.22	88.33	82.56	80.78	91.50	97.44	90.56	95.67	99.22	93.54

### 3 EMSGD 的提出与试验结果

为解决前述不平衡训练数据对 CNN 的不良影响,我们提出改进算法 EMSGD。与随机梯度下降算法的对比试验结果表明,我们提出的改进算法是有效的。

#### 3.1 算法思想与实现

与随机梯度下降算法不同的是,我们的改进算法从原来随机抽取训练数据组成小批量,改进为从每类数据抽取相同数量的样本组成小批量,使得小批量内部保持类别均衡。改进算法的目的是要保证小批量内的数据均衡,因此对少数类,我们采取过抽样,而对多数类,则采取欠抽样。该算法的优点是即使对少数类采取过抽样,只需满足小批量内的均衡即可,不需要像传统的过抽样技术那样满足整个训练数据的均衡,因此少数类的重复样本会比较少。另外,对多数类而言,它只是在小批量内进行了欠采样。由于训练过程是由多个时期(epoch)组成,最终所有的多数类样本都会被用于训练,所以也不存在因欠采样而导致信息丢失的问题。通过保持小批量内的数据均衡,改进算法不仅可以消除数据不平衡带来的不良影响,还可以优化梯度下降的方向,使模型训练收敛更快。

#### 3.2 试验评估

EMSGD 的性能评估采用与前述试验相同的 CNN 结构和数据,激活函数分别采用  $S$  函数、 $T$  函数和  $R$  函数,试验结果见表 6~8,在 LeNet-5 网络的试验结果见表 9。在激活函数为  $S$  函数的情况下,与传统的随机梯度下降算法相比,改进算法的准确率更高。由表 2 可知,数据分布 2 在随机梯度下降算法下测试数据准确率只有 69.94%,而表 6 显示改进算法的测试数据准确率可达 88.74%,提高近 20%。在不同的数据分布下,改进算法都可有效提高少数类的准确率,如分布 2 中 1 类数据的准确率从 67.67%(表 2)提高到 96.22%(表 6)。

在激活函数为  $T$  函数和  $R$  函数的情况下,尽管传统的随机梯度下降算法测试数据准确率可以达到 90%左右,但在某些少数类上准确率仍然较低。如表 3 中分布 3 的 3 类数据的准确率只有 59.56%,而同样情况下的改进算法使少数类的准确率均可达到 80%左右,这说明改进算法可有效提高 CNN 对少数类的分类准确率。

对比表 9 和表 5 可知改进算法在 LeNet-5 同样有效,在数据不平衡的情况下模型准确率有所提升,比如分布 4 的测试数据准确率由 91.77%提高至 94.16%,并且可以有效提高少数类的分类准确率,比如分布 3 的 9 类由 88.78%提高至 95.56%。

表 6 由 EMSGD 训练的激活函数为  $S$  函数的 CNN 网络分类准确率Table 6 Classification accuracy of CNN networks as  $S$  activation function trained by EMSGD %

分布	测试数据 准确率	手写体数字数据准确率										训练数据 准确率
		0	1	2	3	4	5	6	7	8	9	
1	86.52	96.33	93.33	85.11	80.33	80.78	83.33	90.11	87.44	88.44	78.11	95.98
2	88.74	95.89	96.22	85.11	90.22	83.11	85.00	90.89	87.00	80.67	90.44	94.68
3	91.47	96.11	95.56	88.33	89.11	91.78	92.83	92.56	90.11	89.44	87.89	95.35
4	89.79	97.11	97.00	89.78	92.00	90.33	78.00	94.11	90.78	83.78	79.11	96.09

表 7 由 EMSGD 训练的激活函数为  $T$  函数的 CNN 网络分类准确率Table 7 Classification accuracy of CNN networks as  $T$  activation function trained by EMSGD %

分布	测试数据 准确率	手写体数字数据准确率										训练数据 准确率
		0	1	2	3	4	5	6	7	8	9	
1	87.64	96.89	94.11	86.56	82.00	85.67	83.33	92.67	87.78	88.56	77.89	99.99
2	95.08	96.44	98.67	94.22	95.44	95.67	92.00	95.22	94.78	94.89	91.78	99.71
3	94.29	95.89	98.00	93.33	94.33	94.56	90.00	94.89	92.44	93.44	92.33	99.57
4	92.00	98.67	97.44	90.00	93.33	94.67	87.67	94.33	91.22	88.78	79.78	99.69

表 8 由 EMSGD 训练的激活函数为  $R$  函数的 CNN 网络分类准确率Table 8 Classification accuracy of CNN networks as  $R$  activation function trained by EMSGD %

分布	测试数据 准确率	手写体数字数据准确率										训练数据 准确率
		0	1	2	3	4	5	6	7	8	9	
1	88.36	94.67	95.33	89.22	83.44	86.89	83.83	92.11	87.33	89.33	80.44	100.00
2	91.47	94.78	98.67	87.89	91.89	87.67	93.00	91.56	95.11	79.11	93.33	99.95
3	94.57	88.33	98.78	94.78	95.56	96.56	91.17	96.11	93.11	94.56	93.89	99.90
4	92.57	97.89	97.56	93.44	96.33	95.67	91.00	95.78	90.89	86.00	79.33	99.86

表 9 在 LeNet-5 网络 EMSGD 训练的分类准确率

Table 9 Classification accuracy of EMSGD training for LeNet-5 networks %

分布	测试数据 准确率	手写体数字数据准确率										训练数据 准确率
		0	1	2	3	4	5	6	7	8	9	
1	90.42	98.00	95.89	91.11	86.67	88.22	88.00	93.00	89.56	89.78	83.33	100.00
2	92.79	96.33	98.89	89.78	94.11	86.89	96.17	91.67	94.00	83.11	95.44	99.95
3	96.02	95.89	98.22	95.22	96.67	95.44	95.00	96.22	93.56	96.67	95.56	99.84
4	94.16	98.89	97.00	96.33	97.33	94.11	93.67	95.00	91.78	90.44	84.44	99.85

在数据不均衡的情况下,改进算法比随机梯度下降算法可以更快地收敛。图 3~5 分别为激活函数为  $S$  函数、 $T$  函数、 $R$  函数时,CNN 在 4 种不同训练数据分布情况下的训练过程。

由图 3 可知,针对 4 种不同分布的数据,在激活函数为  $S$  函数的情况下,使用改进算法在前期迭代中会更快地达到一个较高的准确率。如图 3(a)所示,改进算法在第 7 500 次迭代时,准确率就达到了 80%,而传统的随机梯度下降算法前期准确率提高较慢,在 7 500 次迭代时准确率只达到 50%。训练完成后,改进算法训练得到的 CNN 测试准确率更高,在数据不均衡的情况下(如图 3(b)、(c)、(d))表现得更为明显。

图 4、图 5 是模型在激活函数为  $T$  函数和  $R$  函数时在 4 种不同数据分布情况下的训练过程。与  $S$  函数的情况相似,改进算法比随机梯度下降算法在不均衡数据集上的收敛速度更快,且准确率更高。但与图 3 不同的是,图 4(a)和图 5(a)中,使用随机梯度下降算法达到的准确率略高于使用改进算法。由对应的表 7 和表 8 可以发现,在分布 1 的情况下,改进算法的训练数据准确率比随机梯度下降算法要高,而测试数据准确率反而降低,由此可推断改进算法可能在数据均衡的情况下易发生过拟合,更适合应用于数据不均衡情况。

图 6 是 LeNet-5 在 4 种不同数据分布情况下的训练过程。与图 5 相似,改进算法比随机梯度下降算法在不均衡数据上的收敛速度更快,且准确率更高,试验证明了我们提出的改进算法在不同的网络结构下均是有效的。但由于 LeNet-5 模型更复杂,易发生过拟合,因此在分布 1 和分布 2 情况下的模型最终的性能没有太大的提升,反而略低于随机梯度下降算法训练得到的模型。

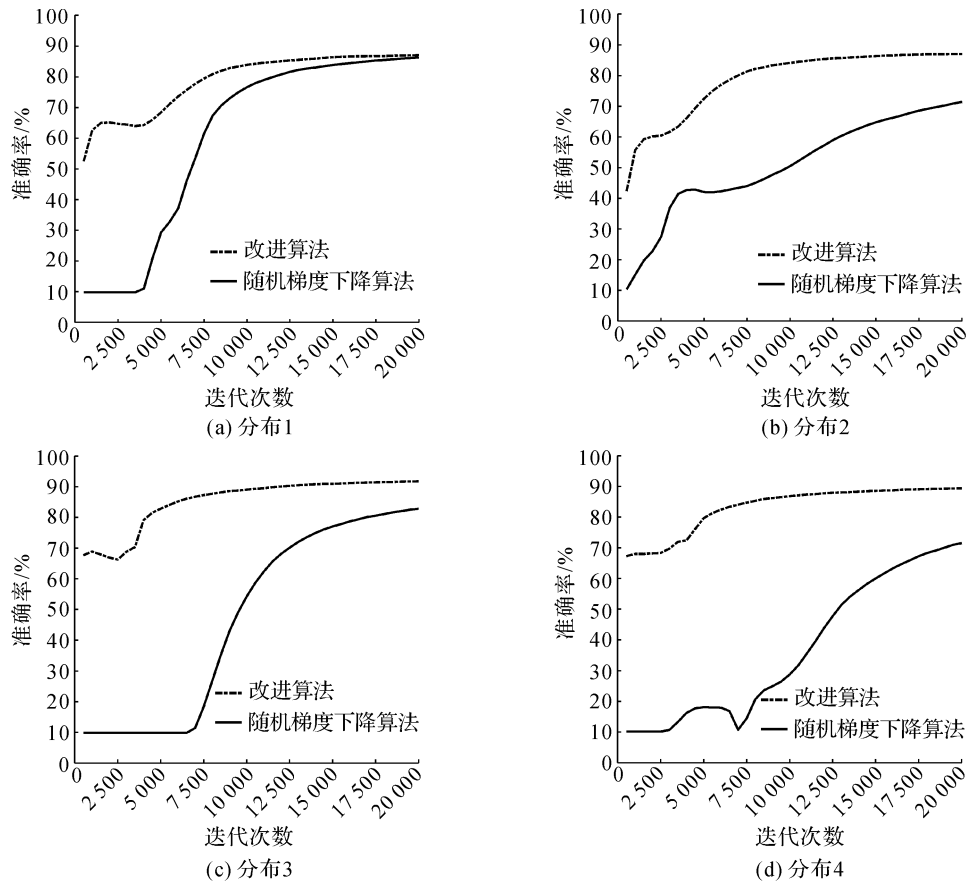


图3 激活函数为S函数的CNN训练过程

Fig. 3 Training process of CNN with S activation function

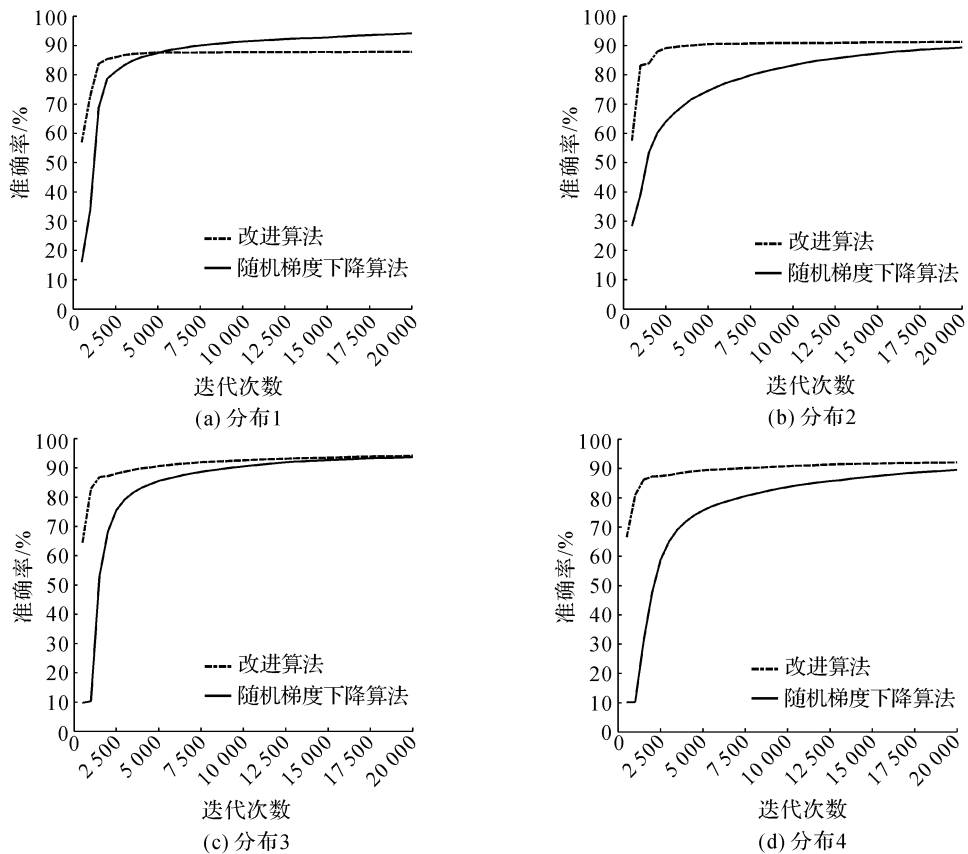


图4 激活函数为T函数的CNN训练过程

Fig. 4 Training process of CNN with T activation function

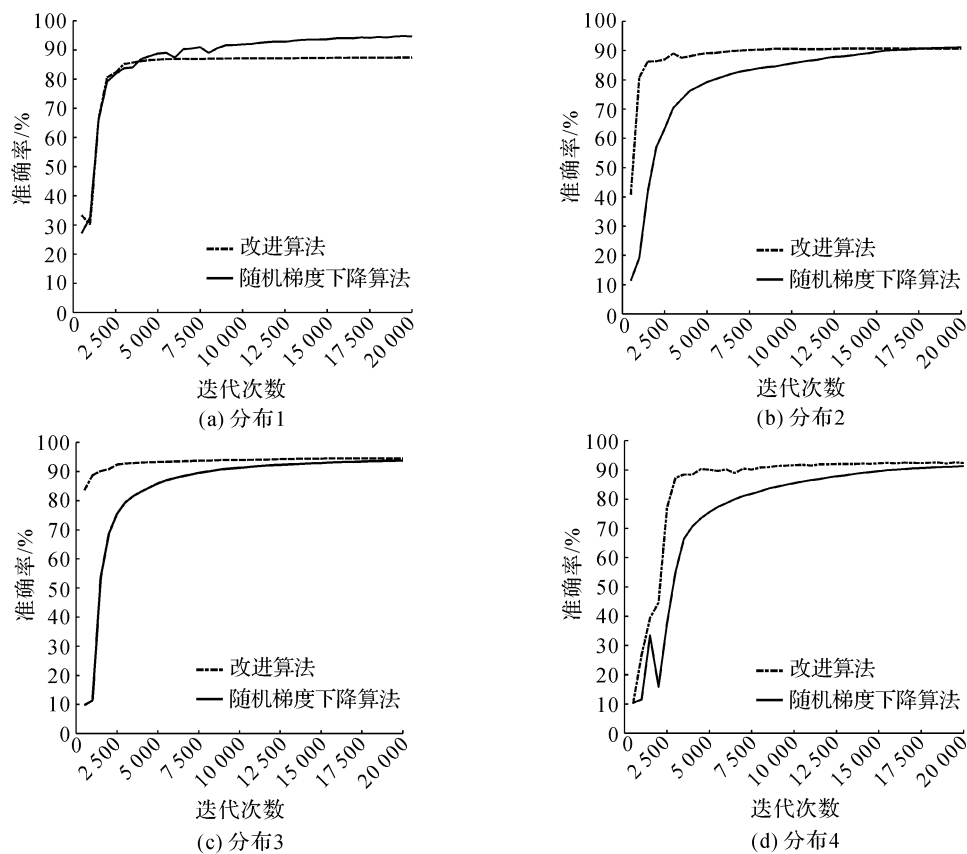


图 5 激活函数为 R 函数的 CNN 训练过程

Fig. 5 Training process of CNN with R activation function

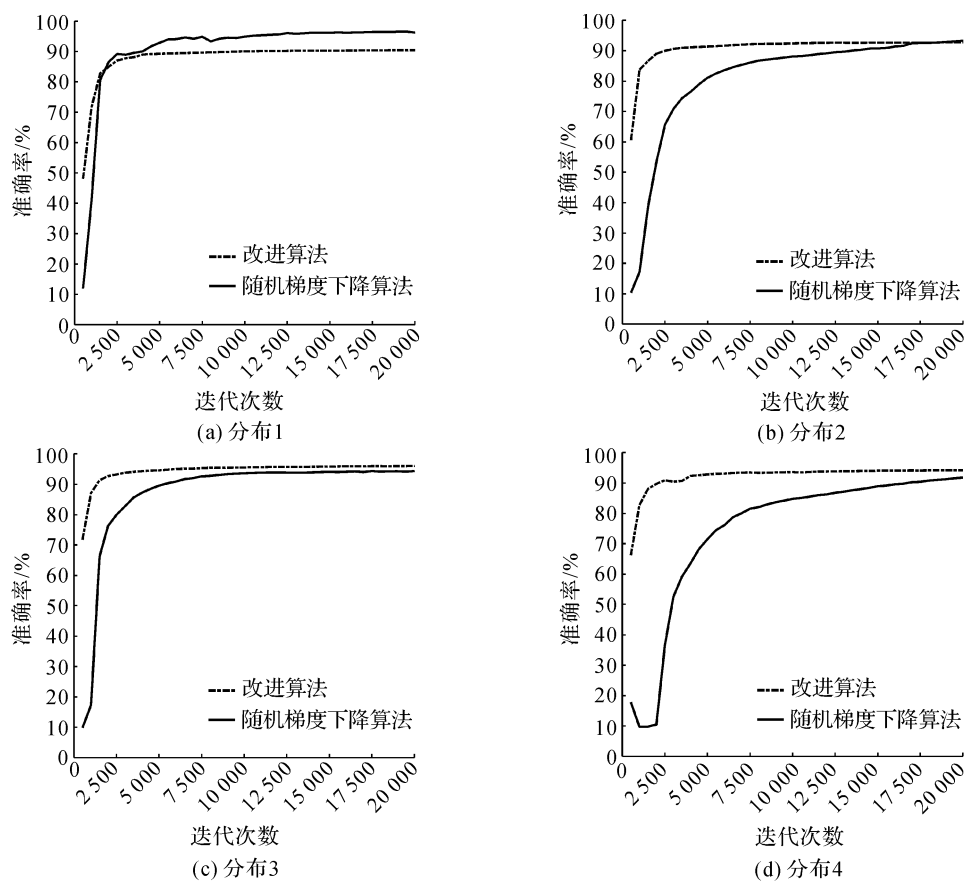


图 6 LeNet-5 的训练过程

Fig. 6 Training process of LeNet-5



## 4 结 语

本文讨论了不平衡数据对 CNN 训练造成的不良影响。通过试验我们发现,在多种数据不平衡情况下 CNN 训练过程中的不良特征,尤其是少数类的准确率较低。试验进一步发现在不同激活函数的情况下都存在这种不良特征,对此我们提出传统的随机梯度下降算法的改进算法——均衡小批量随机梯度下降算法,在每次迭代计算梯度时,保持小批量内数据的类别均衡,克服数据不平衡带来的负面影响。与随机梯度下降算法相比,试验结果表明改进算法在多种激活函数和不同网络结构的情况下,均能提高 CNN 的正确率,并且收敛速度更快,适合在数据不平衡的情况下使用。本文提出的改进方法可以进一步完善,主要方向为如何防止可能会出现的过拟合问题,并可以与 AdaDelta、Adam 等优化方法相结合。

## 参考文献:

- [1] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE,2016:770.
- [2] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks [C]//Advances in Neural Information Processing Systems. Lake Tahoe: NIPS,2012:1097.
- [3] HINTON G, DENG L, YU D, et al. Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups[J]. IEEE Signal Processing Magazine,2012,29(6):82.
- [4] XIONG W, DROPPA J, HUANG X, et al. Toward human parity in conversational speech recognition[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing,2017,25(12):2410.
- [5] SUTSKEVER I, VINYALS O, LE Q V. Sequence to sequence learning with neural networks[C]//Advances in Neural Information Processing Systems. Montreal: NIPS,2014:3104.
- [6] HE D, XIA Y C, QIN T, et al. Dual learning for machine translation[C]//Advances in Neural Information Processing Systems. Baecelona: NIPS,2016:820.
- [7] BHATTACHARYA S, RAJAN V, SHRIVASTAVA H. ICU mortality prediction: a classification algorithm for imbalanced datasets[C]//Proceeding of the Thirty-First AAAI Conference on Artificial Intelligence. San Francisco: AAAI,2017:1288.
- [8] ZAKARYAZAD A, DUMAN E. A profit-driven artificial neural network(ANN) with applications to fraud detection and direct marketing[J]. Neurocomputing,2016,175:121.
- [9] WANG S, YAO X. Using class imbalance learning for software defect prediction [J]. IEEE Transactions on Reliability,2013,62(2):434.
- [10] DUAN L X, XIE M Y, BAI T B, et al. A new support vector data description method for machinery fault diagnosis with unbalanced datasets[J]. Expert Systems with Applications,2016,64:239.
- [11] KHAN S H, HAYAT M, BENNAMOUN M, et al. Cost-sensitive learning of deep feature representations from imbalanced data[J]. IEEE Transactions on Neural Networks and Learning Systems,2015,29(8):3573.
- [12] KRAWCZYK B. Learning from imbalanced data: open challenges and future directions[J]. Progress in Artificial Intelligence,2016,5(4):221.
- [13] GUO H X, LI Y J, JENNIFER S, et al. Learning from class-imbalanced data: review of methods and applications [J]. Expert Systems with Applications,2017,73:220.
- [14] HE H B, GARCIA E A. Learning from imbalanced data [J]. IEEE Transactions on Knowledge and Data Engineering,2009,21(9):1263.
- [15] CHAWLA N V, BOWYER K W, HALL L O, et al. SMOTE: synthetic minority over-sampling technique[J]. Journal of Artificial Intelligence Research,2002,16(1):321.
- [16] DRUMMOND C, HOLTE R C. C4.5, class imbalance, and cost sensitivity: why under-sampling beats over-sampling[C]//Workshop on Learning from Imbalanced Datasets II. Washington DC: Citeseer,2003.
- [17] ELKAN C. The foundations of cost-sensitive learning[C]//International Joint Conference on Artificial Intelligence.

Washington DC: Morgan Kaufmann, 2001.

- [18] MAZUROWSKI M A, HABAS P A, ZURADA J M, et al. Training neural network classifiers for medical decision making: the effects of imbalanced datasets on classification performance[J]. *Neural Networks*, 2008, 21(2/3): 427.
- [19] JACCARD N, ROGERS T W, MORTON E J, et al. Detection of concealed cars in complex cargo X-ray imagery using deep learning[J]. *Journal of X-ray Science and Technology*, 2017, 25(3): 323.
- [20] JANOWCZYK A, MADABHUSHI A. Deep learning for digital pathology image analysis: a comprehensive tutorial with selected use cases[J]. *Journal of Pathology Informatics*, 2016, 7(1): 29.
- [21] CHUNG Y A, LIN H T, YANG S W. Cost-aware pre-training for multiclass cost-sensitive deep learning[C]// *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*. Melbourne: AAAI, 2016: 1411.
- [22] KHAN S H, HAYAT M, BENNAMOUN M, et al. Cost-sensitive learning of deep feature representations from imbalanced data[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2015, 29(8): 3573.
- [23] HAVAEI M, DAVY A, WARDE-FARLEY D, et al. Brain tumor segmentation with deep neural networks[J]. *Medical Image Analysis*, 2017, 35: 18.
- [24] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. *Proceedings of the IEEE*, 1998, 86(11): 2278.

~~~~~

(上接第 173 页)

- [9] GHOSH A K, CHAPSAL B D, MITSUYA H. Design of HIV protease inhibitors targeting protein backbone: an effective strategy for combating drug resistance[J]. *Accounts of Chemical Research*, 2008, 41(1): 78.
- [10] SIMEONOV S P, NUNES J P M, GUERRA K, et al. Synthesis of chiral cyclopentenones[J]. *Chemical Reviews*, 2016, 116(10): 5744.
- [11] AITKEN D J, EIJSBERG H, FRONGIA A, et al. Recent progress in the synthetic assembly of 2-cyclopentenones[J]. *Synthesis*, 2014, 46(1): 1.
- [12] EVONNE A B, JILLIAN M B. Optical sensing of amine vapors with a series of tin compounds[J]. *Chemical Communications*, 2004, 2(18): 2028.
- [13] NICOLAOU K C, PHILIPP H, ABDELATIF E, et al. 18. Total synthesis of  $\Delta^{12}$ -prostaglandin  $J_3$ , a highly potent and selective antileukemic agent[J]. *Angewandte Chemie*, 2014, 53(39): 10443.
- [14] SHIMADA N, STEWART C, TIUS M A. Asymmetric nazarov cyclizations[J]. *Tetrahedron*, 2011, 67(33): 5851.
- [15] YAMAGUCHI M, SHIBATO K, FUJIWARA S, et al. A convenient synthesis of alkynyl ketones from esters[J]. *Synthesis*, 1986(5): 421.