

DOI: 10.11992/tis.201804056

网络出版地址: <http://kns.cnki.net/kcms/detail/23.1538.TP.20180610.2155.002.html>

基于改进卷积神经网络的多标记分类算法

余鹰¹, 王乐为¹, 吴新念¹, 伍国华², 张远健³

(1. 华东交通大学软件学院, 江西南昌 330013; 2. 中南大学交通运输工程学院, 湖南长沙 410000; 3. 同济大学计算机科学与技术系, 上海 201804)

摘要: 良好的特征表达是提高模型性能的关键, 然而当前在多标记学习领域, 特征表达依然采用人工设计的方式, 所提取的特征抽象程度不高, 包含的可区分性信息不足。针对此问题, 提出了基于卷积神经网络的多标记分类模型 ML_DCCNN, 该模型利用卷积神经网络强大的特征提取能力, 自动学习能刻画数据本质的特征。为了解决深度卷积神经网络预测精度高, 但训练时间复杂度不低的问题, ML_DCCNN 利用迁移学习方法缩减模型的训练时间, 同时改进卷积神经网络的全连接层, 提出双通道神经元, 减少全连接层的参数量。实验表明, 与传统的多标记分类算法以及已有的基于深度学习的多标记分类模型相比, ML_DCCNN 保持了较高的分类精度并有效地提高了分类效率, 具有一定的理论与实际价值。

关键词: 多标记学习; 卷积神经网络; 迁移学习; 全连接层; 特征表达; 多标记分类; 深度学习; 损失函数

中图分类号: TP181 **文献标志码:** A **文章编号:** 1673-4785(2019)03-0566-09

中文引用格式: 余鹰, 王乐为, 吴新念, 等. 基于改进卷积神经网络的多标记分类算法[J]. 智能系统学报, 2019, 14(3): 566-574.

英文引用格式: YU Ying, WANG Lewei, WU Xinnian, et al. A multi-label classification algorithm based on an improved convolutional neural network[J]. CAAI transactions on intelligent systems, 2019, 14(3): 566-574.

A multi-label classification algorithm based on an improved convolutional neural network

YU Ying¹, WANG Lewei¹, WU Xinnian¹, WU Guohua², ZHANG Yuanjian³

(1. College of Software Engineering, East China Jiaotong University, Nanchang 330013, China; 2. College of Transportation Engineering, Central South University, Changsha 410000, China; 3. Department of Computer Science and Technology, Tongji University, Shanghai 201804, China)

Abstract: A good feature expression is the key to improve model performance. However, at present, artificially designed features are used for multi-label learning. Thus, the level of abstraction of the extracted features is low and lacks the discriminated information involved. To solve this problem, this paper proposes a multi-label classification model based on convolutional neural network (ML_DCCNN). This model uses the powerful feature extraction capabilities of CNNs to automatically learn the features from the data. To solve the problem of high forecasting precision versus long training time of CNNs, the ML_DCCNN uses the transfer learning method to reduce the training time of the model. In addition, the entire connection layer of the CNN is improved by a dual-channel neuron, which can reduce the number of parameters of the fully connected layer. The experiments show that compared with the traditional multi-label classification algorithm and existing multi-label classification model based on deep learning, the ML_DCCNN maintains high classification accuracy and can effectively improve the classification efficiency, presenting certain theoretical and practical value.

Keywords: multi-label learning; convolutional neural network; transfer learning; fully-connected layer; feature expression; multi-label classification; deep learning; loss function

收稿日期: 2018-04-26. 网络出版日期: 2018-06-11.

基金项目: 国家自然科学基金项目 (61563016, 61603404, 61462037, 61663002); 江西省教育厅科技项目 (GJJ150546); 江西省自然科学基金项目 (2018BAB202023).

通信作者: 余鹰. E-mail: yuyingjx@163.com.

不同于传统的单标记学习问题, 多标记学习考虑一个对象对应多个类别标记的情况。例如: 一个基因可能同时具有多种功能, 如新陈代谢、

转录以及蛋白质合成;一首乐曲可能传达了多种信息,如钢琴、古典音乐和莫扎特等;一幅图像可能同时属于多个类别,如 motor、person 与 car 等。早期,多标记学习的研究主要集中于文本分类中遇到的多义性问题。经过近十年的发展,多标记学习已成为当前国际机器学习领域研究的热点问题之一,逐渐在情感分类^[1]、图像视频语义标注^[2]、生物信息学^[3]和个性化推荐^[4]等实际应用中扮演重要的角色。随着相关应用的发展及需求的不断提升,多标记学习技术的大规模应用仍然要应对很多的问题和挑战。当前在多标记学习领域,特征表达大多采用人工设计的方式,如 SIFT、HOG 等,这些特征在特定类型对象中能够达到较好的识别效果,但这些算法提取的只是一些低层次(low-level)特征,抽象程度不高,包含的可区分性信息不足,对于分类来说无法提供更多有价值的语义信息,影响分类的精度。目前,如何让多标记系统学会辨别底层数据中隐含的区分性因素,自动学习更抽象和有效的特征已成为制约多标记学习研究进一步深入的瓶颈。

近年来,深度学习在图像分类和目标检测等领域取得了突破性进展,成为目前最有效的特征自动学习方法。文献[5]将传统人工设计的特征与深度神经网络自学习的特征进行了比较,发现后者更有助于提升图像自动标注算法的性能。深度学习模型具有强大的表征和建模能力,通过监督或非监督的方式,逐层自动地学习目标的特征表示,将原始数据经过一系列非线性变换,生成高层次的抽象表示,避免了手工设计特征的烦琐低效。本文针对多标记学习中存在的特征抽象层次不高的问题,利用包含多个隐含层的深度卷积神经网络直接从原始输入中学习并构建多层的分级特征,形成更加抽象的高层表示,实现以最少和最有效的特征来表达原始信息。同时,针对卷积神经网络预测精度高但运算速度慢的特点,利用迁移学习和双通道神经元方法,缩减网络的参数量,提高训练速度,在一定程度上弥补了卷积神经网络计算量大、速度较慢的缺陷。

1 相关工作

1.1 多标记学习

为了便于叙述,在分析之前先给出多标记问题的形式化定义。令 $X = \{x_1, x_2, \dots, x_m\}$ 代表实例空间, $L = \{l_1, l_2, \dots, l_q\}$ 代表所有标记的集合, $Y = \{y_1, y_2, \dots, y_m\}$ 代表标记空间, $T = \{(x_i, y_i) | 1 \leq i \leq m\}$ 代表训练集。多标记分类的任务就是用训练集 T 对

模型进行训练,使模型学得函数 $f: X \rightarrow Y$, 其中 $x_i \in X$ 表示一个实例, $y_i \in Y$ 表示实例 x_i 所含有的类别标记,且 y_i 为标记集合 L 的一个子集。训练完成后,将未分类的数据输入模型,得到与实际标记最大程度接近的分类结果。

目前,多标记分类算法根据解决问题方式的不同,可归为问题转换型和算法适应型两类^[6]。问题转换型是将多标记分类问题转化为多个单标记分类问题,如算法 BR(binary relevance)^[7]、LP(label powerset)^[8]等,然后利用单标记分类方法进行处理。算法适应型则是改进已有的单标记分类算法,使其适应于多标记分类问题,如算法 BSVM(biased support vector machine)^[9]、ML-KNN(multi-label k-nearest neighbor)^[10]等。随着深度学习的兴起,已有一些学者开始基于深度学习研究多标记分类问题,Zhang^[11]由传统径向基函数 RBF(radial basis function)推导出了一种基于神经网络的多标记学习算法 ML-RBF。Wang 等^[12]将卷积神经网络 CNN(convolutional neural network)和循环神经网络 RNN(recurrent neuron network)相结合,提出了一种多标记学习的复合型框架,用于解决多标记图像分类问题,但这些算法的精度和时间复杂度都有待进一步提升。

1.2 卷积神经网络

卷积神经网络 CNN 是一种深度神经网络模型,主要由卷积层、池化层和全连接层构成,如图 1 所示。卷积层负责图像特征提取,池化层用于降维及实现不变形,而全连接层则起到分类器的作用。卷积层和池化层一般作为组合多次成对出现,也可以根据实际情况灵活使用,如 AlexNet^[13]和 VGG^[14]。



图 1 卷积神经网络结构

Fig. 1 Convolutional neural network structure

相比于传统的特征提取方法,卷积神经网络不需要事先人工设定特征,而是通过网络模型从大量数据中自动学习特征表示。通过多层非线性映射,逐层提取信息,最底层从像素级原始数据学习滤波器,刻画局部边缘和纹理特征;中层滤波器对各种边缘滤波器进行组合后,描述不同类型的局部特征;最高层描述整体全局特征。

1.3 迁移学习

迁移学习 (transfer learning) 的基本思想是将

从一个环境中学到的知识用于新环境中的学习任务。

目前,迁移学习已被广泛应用于各个领域,例如,在文档分类方面,Dai等^[15]提出联合聚类的方法,通过不同领域共享相同的词特征进行知识迁移;在智能规划中,Zhuo等^[16]提出一种新的迁移学习框架 TRAMP,通过建立源领域与目标领域之间的结构映射来迁移知识,获取人工智能规划中的动作模型。

2 基于改进 CNN 的多标记分类算法

2.1 算法框架

由于图像传递信息的底层机制相通,因此可以利用迁移学习,将在源域上训练好的网络模型,通过共享网络参数,使之在目标域上也具有一定的特征提取能力。本文采用在 ImageNet^[17]数据集上训练好的 Inception V3^[18]模型进行图像特征提取。该模型引入了“Factorization into small convolutions”的思想,将一个较大的二维卷积核

拆分成两个较小的一维卷积核,例如将 3×3 的卷积核分解成 1×3 和 3×1 两个卷积核。这种非对称的拆分方式减少了网络参数量,降低了过拟合风险,可以保证对 ImageNet 之外的数据集具有很好的泛化性能。

为了进一步减少全连接层参数数量,本文对 Inception V3 模型的全连接层进行改进,引入双通道神经元,优化网络结构,并结合迁移学习提出了多标记分类模型 ML_DCCNN。最后,将全连接层的输出送入 SoftMax 分类器,从而得到各标记的预测概率,然后根据各标记的概率计算多标记分类损失函数。

在反向传播时,保留 Inception V3 模型的特征提取层,即固定特征提取层的权重和偏置参数,并用神经元个数为 20 的全连接层替换原有全连接层,设置该层的初始权重和偏置为 0,学习率设置为 0.001, batchsize 设置为 100。然后,使用随机梯度下降算法,用 PASCAL Visual Object Classes Challenge(VOC)数据集^[19]对网络参数进行微调,使其适应于新数据集,算法的具体流程如图 2 所示。

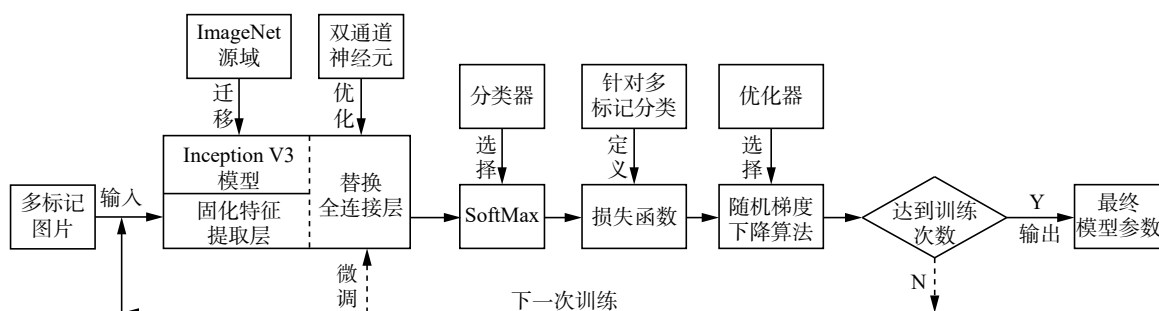


图 2 基于改进 CNN 的多标记分类算法框架

Fig. 2 Multi-label classification algorithm framework based on improved convolution neural network

2.2 双通道神经元

在卷积神经网络中,卷积、池化和激活函数等操作将原始数据映射到隐层特征空间,全连接层则将学到的分布式特征表示映射到标记空间,即全连接层在整个卷积神经网络中起到了“分类器”的作用。但全连接层上往往包含大量参数,对整个网络的速度有一定影响。虽然 FCN^[20]全卷积模型取消了全连接层,避免了全连接层的副作用,但是在 Zhang 等^[21]的研究中,全连接层能够在模型表示能力迁移过程中充当“防火墙”的作用,保证模型表示能力的迁移。因此为了能够在保留全连接层的基础上,减少网络参数,本文提出了双通道神经元的概念。

2.2.1 基本结构

本文将全连接层中只能接受一种标记特征信

息的神经元称为普通神经元,如图 3(a) 所示,全连接层中最后一层的神经元个数和具体分类问题的标记总数相等,如某数据集上共有 n 种标记,则最后一层全连接层上的神经元个数为 n 。

本文将可以接受两种标记特征信息的神经元称为双通道神经元。一个双通道神经元相当于两个普通神经元的合并,它改进了全连接层,有效地减少了该层的参数。在接受到特征信息后,为了能将合并的标记区分,使用双通道的神经元,需在其后再连接两个神经元,分别表示对应的标记,并规定所连接的两个神经元上的权重分别为 1 和 -1,如图 3(b) 所示。

在图 3(a) 中,假设分类标记数为 n ,全连层输入为 m ,则权重参数有 $m \times n$ 个,偏置参数有 n 个,总参数有 $m \times n + n$ 个。

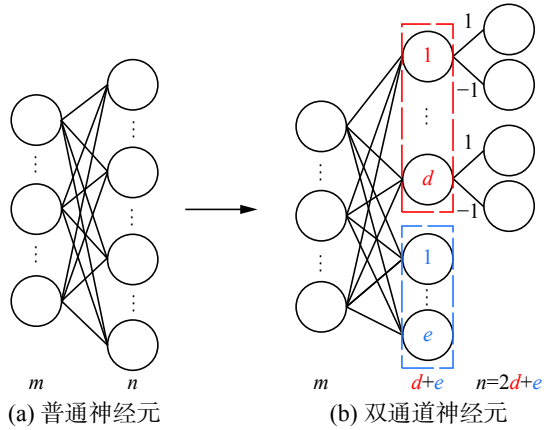


图3 全连接层

Fig. 3 Fully connected layer

在图3(b)中,假设全连接层有 d 个双通道神经元和 e 个不使用双通道的神经元。在同样假设条件下,该层权重参数为 $m \times (d+e) + 2d$ 个,偏置参数为 $d+e$ 个,总参数为 $(m+1) \times (d+e) + 2d$ 个。一般情况下,输入值 m 远大于输出值 n ,例如 In-

ception V3 模型的输入 m 为2048,输出标记 n 为20,所以在 m 远大于 n 前提下,使用双通道神经元最多可缩减一半参数,如公式(1)所示:

$$\frac{1}{2} \leq \frac{(m+1)(d+e)+2d}{(m+1)n} \leq 1 \quad (1)$$

式中: $m \gg n; n = 2d + e$ 。

2.2.2 核心理想

打包和解包是双通道神经元的核心理想。打包主要表现在将两种标记合二为一在一个神经元上,即最后一层全连接层上的每个神经元可以表示两种标记,接受两种标记的特征信息。例如:将飞机和自行车这两种标签打包在一起,由一个神经元负责输出,则该神经元上的权重只对飞机和自行车的特征信息敏感。但仅用一个神经元输出,存在无法判别输出是飞机还是自行车的情况,因此需要解包思想,主要表现在一个神经元又“分裂”出两个神经元,具体如图4所示。

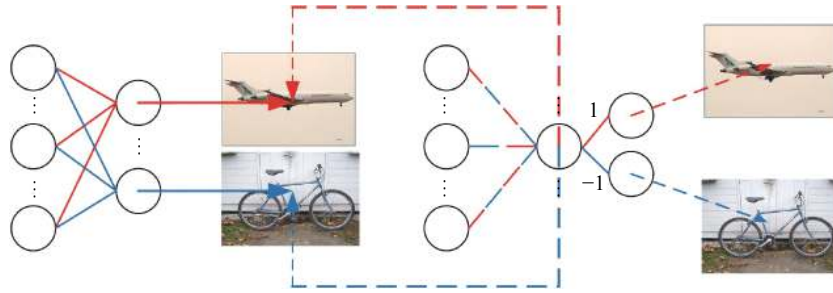


图4 打包与解包示意

Fig. 4 Package and unpack diagram

图4左边为普通全连接层的神经元,每个神经元仅对一种标记特征信息敏感,如上方神经元仅对飞机特征信息敏感,下方神经元仅对自行车特征信息敏感。图4右边使用了双通道神经元,每个神经元对两种类别的特征信息敏感,例如同时对飞机和自行车的特征信息敏感,在提取出飞机和自行车的特征后,再分裂出两个神经元分别代表对应的标记,其中权重为1的代表飞机,权重为-1的代表自行车。

2.3 损失函数

设 $D = \{(\mathbf{x}_i, \mathbf{y}_i) | i = 1, 2, \dots, n\}$ 代表具有 n 个样本的训练集,其中 $\mathbf{x}_i = [x_{i1} \ x_{i2} \ \dots \ x_{id}]$ 是第 i 个样本的 d 维特征向量, $\mathbf{y}_i = [y_{i1} \ y_{i2} \ \dots \ y_{iq}]$ 是第 i 个样本的标记向量,其维度 q 与数据集标记总数相等, $y_{ij} = 1$ 表示 \mathbf{x}_i 含有标签 l_j , $y_{ij} = 0$ 则表示不含有。

SoftMax 分类器不仅可以用于处理单标记分类问题,也可以用于处理多标记分类问题。本文将最后一层全连接层的输出送入 SoftMax 分类器

中,得出图片含有各标记的概率,例如图片 \mathbf{x}_i 含有标记 l_j 的概率:

$$p_{ij} = \frac{\exp(f_j(\mathbf{x}_i))}{\sum_{k=1}^c \exp(f_k(\mathbf{x}_i))} \quad (2)$$

式中: $f_j(\mathbf{x}_i)$ 表示图片 \mathbf{x}_i 对应标记 l_j 的激活值, q 表示数据集的标记总数。经过 SoftMax 分类器输出各标记概率后,定义交叉熵损失函数:

$$J = - \sum_{i=1}^n \sum_{j=1}^q \overline{p}_{ij} \log(p_{ij}) \quad (3)$$

式中: \overline{p}_{ij} 表示图片 \mathbf{x}_i 是否含有标记 j ;当图片 \mathbf{x}_i 含有 j 标记时, \overline{p}_{ij} 值为 $\frac{1}{c_+}$,当图片 \mathbf{x}_i 不含有标记 j 时, \overline{p}_{ij} 值为0,如式(4):

$$\overline{p}_{ij} = \begin{cases} \frac{1}{c_+}, & y_{ij} = 1 \\ 0, & y_{ij} = 0 \end{cases} \quad (4)$$

由式(3)和式(4)可以推导出:

$$J = - \sum_{i=1}^n \sum_{j=1}^{c_+} \frac{1}{c_+} \log(p_{ij}) \quad (5)$$

式中: n 表示一个 Batch 上的图片数量; c_i 表示图片 x_i 上正标记的总个数。

3 实验与分析

本文实验在处理器为 i5-3210M 的 Windows PC 机上完成, 基于 TensorFlow 1.2.1 实现卷积神经网络, 采用了 PASCAL VOC2007 和 PASCAL VOC2012 两个多标记数据集, 二者均含有 20 个类别标记。PASCAL VOC2007 数据集共有 9 963 张图片, 其中训练验证集有 5 011 张, 测试集有 4 952 张, PASCAL VOC2012 共有 33 260 张图片, 其中训练验证集有 17 125 张, 测试集有 16 135 张。

为了验证双通道神经元的可用性, 本文对普通全连接层结构和采用双通道神经元的全连接层的分类效果进行了比较, 其中双通道神经元的标签两两合并方式如表 1 所示。表 2 显示了 PASCAL VOC 数据集中的一个多标记图像(如图 5) 分别使用普通全连接层和包含 10 个双通道神经元的全连接层在训练 2 000 步时, softmax 分类器的输出值, 其中, FC 表示普通全连接层, DC(Dual_Channel) 表示双通道神经元全连接层, GT 表示 ground_truth。DC 所用 d 值为 10, 即 10 个双通道神经元结构。表 3 为训练了 5 000 次的普通全连接层 FC 和使用 10 个双通道神经元的全连接层 DC 在 100 张图片上每个标记 softmax 的平均输出值, 其中 NUM 表示在 100 张图片上每个标记出现的总数, 出现次数最多的标记分别为 person 和 car, 可以看出两种全连接层在这两个标记上所对应的 softmax 也最大, 且很接近。由表 2 和表 3 可知, 在使用 10 个双通道神经元时, 全连接层参数缩减了近一半, 但网络仍能准确识别出多标记图片中的物体类别, 且 softmax 分类器的输出值与普通全连接层相近, 在一定程度上证明了双通道神经元的可用性。



图 5 多标记图像

Fig. 5 Multi-label image

表 1 标签合并方式
Table 1 Label merging mode

序号	标签 1	标签 2
1	plane	bike
2	bird	boat
3	bottle	bus
4	car	cat
5	chair	cow
6	table	dog
7	horse	motor
8	person	plant
9	sheep	sofa
10	train	TV

表 2 使用两种全连接层的分类结果比较
Table 2 Result comparison of two fully connected layers

标签	FC	DC	GT
motor	0.312 7	0.337 4	1
person	0.206 4	0.183 9	1
car	0.096 7	0.078 9	1
dog	0.037 7	0.043 0	0
bird	0.031 1	0.037 0	0
cat	0.030 1	0.008 9	0
bike	0.027 5	0.031 6	0
plant	0.024 7	0.003 8	0
train	0.024 5	0.034 3	0
chair	0.023 8	0.032 8	0
plane	0.021 9	0.022 3	0
bus	0.021 5	0.030 0	0
bottle	0.021 2	0.023 5	0
horse	0.020 6	0.002 1	0
TV	0.019 7	0.020 5	0
sofa	0.018 0	0.025 7	0
boat	0.016 5	0.019 0	0
sheep	0.016 1	0.027 4	0
table	0.014 9	0.016 4	0
cow	0.014 4	0.021 5	0

表3 两种全连接层的平均分类效果比较

Table 3 Average effect comparison of two fully connected layers

标签	FC	DC	NUM
plane	0.040 1	0.050 1	5
bike	0.051 2	0.057 4	8
bird	0.060 8	0.055 0	6
boat	0.031 5	0.022 4	5
bottle	0.027 1	0.028 9	4
bus	0.027 0	0.026 6	2
car	0.130 6	0.131 2	15
cat	0.038 5	0.035 9	4
chair	0.045 6	0.047 7	9
cow	0.017 7	0.020 5	1
table	0.020 0	0.024 0	5
dog	0.051 5	0.051 2	7
horse	0.023 1	0.026 5	4
motor	0.046 2	0.053 4	6
person	0.253 4	0.234 6	47
plant	0.033 6	0.009 5	2
sheep	0.027 3	0.030 8	4
sofa	0.023 3	0.029 0	4
train	0.029 5	0.036 8	4
TV	0.022 0	0.028 5	1

为了说明双通道神经元个数对分类效果的影响,在PASCAL VOC2007数据集上对比了双通道神经元个数 d 分别取不同值时的模型分类效果,如表4所示。由表3可见,随着 d 值增大,平均准确率(mean average precision, mAP)值在降低,每个标记所得最高AP值出现在非双通道神经元的输出中,即双通道神经元在减少参数的同时牺牲一定的准确率,变化趋势如图6所示。图6中, λ 表示双通道神经元所占的比例, $\lambda=d/n$ 。由图可见,随着 λ 增大,mAP逐渐降低,当 $\lambda=0.5$ 时,mAP取最小值;实线表示全连接层上的参数缩减比例 β ,随着 λ 增加, β 逐步上升,当 $\lambda=0.5$ 时, β 达到最大值0.5。

为了验证ML_DCCNN模型的分类效果,本文分别在Pascal VOC2007和Pascal VOC2012数据集上进行实验,比较了ML_DCCNN、普通全连接层模型CNN-SoftMax、传统的多标记分类算法INRIA^[22]、FV^[23]和GS-MKL^[24],以及基于卷积神经网络的多标记分类模型PRE-1000C^[25]和HCP-1000^[26],评价指标使用Accuracy Precision(AP),双

通道神经元个数 d 取10,实验结果如表5和表6所示。由表可见,相对于传统算法的多标记分类算法,深度神经网络在图像特征提取和分类效果上有着明显优势,其中PRE-1000C和HCP-1000基于AlexNet卷积神经网络,相比于本文迁移学习所使用的Inception V3模型所得mAP值较低。

表4 d 取不同值时AP在PASCAL VOC2007数据集上的变化

Table 4 Comparison of AP of algorithm on PASCAL VOC2007 data set

标签	$d=10$	$d=8$	$d=5$	$d=2$
plane	0.941	0.945	0.953	0.947
bike	0.905	0.911	0.910	0.905
bird	0.903	0.897	0.909	0.913
boat	0.895	0.892	0.906	0.905
bottle	0.685	0.677	0.682	0.708
bus	0.913	0.906	0.918	0.940
car	0.938	0.929	0.925	0.952
cat	0.902	0.906	0.913	0.939
chair	0.576	0.565	0.569	0.602
cow	0.771	0.762	0.767	0.795
table	0.747	0.756	0.785	0.781
dog	0.924	0.917	0.955	0.943
horse	0.895	0.903	0.940	0.947
motor	0.887	0.879	0.919	0.926
person	0.935	0.941	0.958	0.952
plant	0.652	0.665	0.709	0.712
sheep	0.838	0.862	0.889	0.886
sofa	0.702	0.797	0.783	0.793
train	0.933	0.960	0.955	0.957
TV	0.781	0.808	0.816	0.811
mAP	0.836	0.844	0.858	0.866

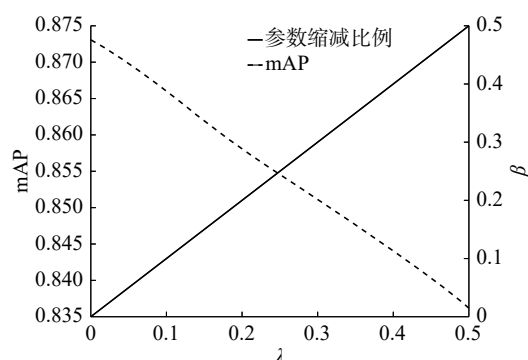
图6 双通道神经元比例 λ 对平均准确率mAP和参数缩减比例 β 的影响Fig. 6 Effect of dual-channel neuron ratio λ on the mAP and ratio β of parameter reduction

表5 不同分类算法 AP 在 PASCAL VOC2007 上的比较

Table 5 Comparison of AP of different classification algorithms on PASCAL VOC2007 data set

标签	INRIA	FV	GS-MKL	PRE-1000C	HCP-1000	CNN-SoftMax	ML_DCCNN
plane	0.772	0.757	0.794	0.885	0.951	0.965	0.941
bike	0.693	0.648	0.624	0.815	0.901	0.932	0.905
bird	0.562	0.528	0.585	0.879	0.928	0.942	0.903
boat	0.666	0.706	0.702	0.820	0.899	0.927	0.895
bottle	0.455	0.300	0.466	0.475	0.515	0.703	0.685
bus	0.681	0.641	0.623	0.755	0.800	0.956	0.913
car	0.834	0.775	0.756	0.901	0.917	0.955	0.938
cat	0.536	0.555	0.549	0.872	0.916	0.943	0.902
chair	0.583	0.556	0.638	0.616	0.577	0.609	0.576
cow	0.511	0.418	0.407	0.757	0.778	0.805	0.771
table	0.622	0.563	0.583	0.673	0.709	0.781	0.747
dog	0.452	0.417	0.516	0.855	0.893	0.955	0.924
horse	0.784	0.763	0.792	0.835	0.893	0.945	0.895
motor	0.697	0.644	0.681	0.800	0.854	0.932	0.887
person	0.861	0.827	0.871	0.956	0.930	0.952	0.935
plant	0.524	0.283	0.495	0.608	0.640	0.704	0.652
sheep	0.544	0.397	0.488	0.768	0.857	0.891	0.838
sofa	0.543	0.566	0.564	0.580	0.627	0.787	0.702
train	0.758	0.797	0.759	0.904	0.944	0.962	0.933
TV	0.621	0.515	0.544	0.779	0.783	0.803	0.781
mAP	0.635	0.583	0.622	0.777	0.815	0.873	0.836

表6 不同分类算法 AP 在 PASCAL VOC2012 数据集上的比较

Table 6 Comparison of AP of different classification algorithms on PASCAL VOC2012 data set

标签	PRE-1000C	PRE-1512	HCP-1000	CNN-SoftMax	ML_DCCNN
plane	0.935	0.946	0.977	0.982	0.968
bike	0.784	0.829	0.830	0.941	0.913
bird	0.877	0.882	0.932	0.927	0.895
boat	0.809	0.841	0.872	0.930	0.904
bottle	0.573	0.603	0.596	0.695	0.667
bus	0.850	0.890	0.882	0.933	0.906
car	0.816	0.844	0.819	0.929	0.895
cat	0.894	0.907	0.947	0.952	0.922
chair	0.669	0.721	0.669	0.706	0.664
cow	0.738	0.868	0.816	0.885	0.853
table	0.620	0.690	0.680	0.803	0.753
dog	0.895	0.921	0.930	0.920	0.911
horse	0.832	0.934	0.882	0.935	0.893
motor	0.876	0.886	0.877	0.932	0.898
person	0.958	0.961	0.927	0.956	0.936
plant	0.614	0.643	0.590	0.683	0.645
sheep	0.790	0.866	0.851	0.892	0.855
sofa	0.543	0.623	0.554	0.736	0.697
train	0.880	0.911	0.930	0.953	0.926
TV	0.783	0.798	0.772	0.828	0.812
mAP	0.787	0.828	0.817	0.876	0.846

总之,通过实验可以证明使用双通道神经元能够对全连接层参数进行一定比例的缩减,而由于全连接层参数往往是迁移学习过程中所需要训练的全部参数,因此全连接层参数的缩减在一定程度上意味着整个网络模型的参数缩减。虽然双通道神经元在特征提取方面存在一定准确率损失,但整体性能依然在可接受范围之内,双通道神经元提供了不同程度的参数缩减与性能表现的可选择性,某种程度上增加了网络模型的灵活性。

4 结束语

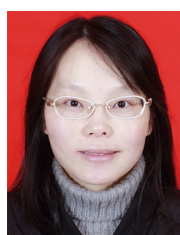
本文提出了一种基于卷积神经网络的多标记分类方法,设定了针对多标记分类的损失函数,并在 PASCAL VOC2007 和 PASCAL VOC2012 两个多标记数据集上进行了验证。总体而言,与以往的方法相比,本文提出的使用迁移学习和双通道神经元多标记分类方法,可以在保证一定准确率的前提下减少网络参数,节省计算资源。在当下注重准确率和计算量平衡的背景下,有着较好的适应性和应用前景。但限于数据、机器性能等因素,本文没有进行更多的实验来证明标记相关性约束条件下分类算法的性能。因此将来的工作从以下方面开展:利用深度学习模型构建标记之间的依赖关系以及在标记依赖关系约束下进行多标记卷积神经网络的训练。

参考文献:

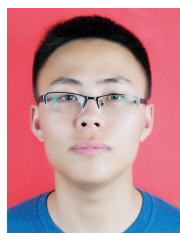
- [1] TROHIDIS K, TSOUMAKAS G, KALLIRIS G, et al. Multilabel classification of music into emotions[C]//Proceedings of 2008 International Conference on Music Information Retrieval (ISMIR 2008). Philadelphia, USA, 2008: 325–330.
- [2] WU Baoyuan, LYU S, HU Baoguang, et al. Multi-label learning with missing labels for image annotation and facial action unit recognition[J]. *Pattern recognition*, 2015, 48(7): 2279–2289.
- [3] JIANG J Q, MCQUAY L J. Predicting protein function by multi-label correlated semi-supervised learning[J]. *IEEE/ACM transactions on computational biology and bioinformatics*, 2012, 9(4): 1059–1069.
- [4] OZONAT K, YOUNG D. Towards a universal marketplace over the web: statistical multi-label classification of service provider forms with simulated annealing[C]//Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Paris, France, 2009: 1295–1304.
- [5] GONG Y, JIA Y, LEUNG T, et al. Deep convolutional ranking for multilabel image annotation[C]. 2nd International Conference on Learning Representations, ICLR 2014. Banff, Canada, 2014: 1312–1320.
- [6] ZHANG Minling, ZHOU Zhihua. A review on multi-label learning algorithms[J]. *IEEE transactions on knowledge and data engineering*, 2014, 26(8): 1819–1837.
- [7] LUACES O, DÍEZ J, BARRANQUERO J, et al. Binary relevance efficacy for multilabel classification[J]. *Progress in artificial intelligence*, 2012, 1(4): 303–313.
- [8] READ J, PFAHRINGER B, HOLMES G. Multi-label classification using ensembles of pruned sets[C]//ICDM'08. Eighth IEEE International Conference on Data Mining. Pisa, Italy, 2008: 995–1000.
- [9] WAN Shupeng, XU Jianhua. A multi-label classification algorithm based on triple class support vector machine[C]//Proceedings of 2007 International Conference on Wavelet Analysis and Pattern Recognition. Beijing, China, 2008: 1447–1452.
- [10] 张敏灵. 一种新型多标记懒惰学习算法[J]. 计算机研究与发展, 2012, 49(11): 2271–2282.
- [11] ZHANG Minling. An improved multi-label lazy learning approach[J]. *Journal of computer research and development*, 2012, 49(11): 2271–2282.
- [12] ZHANG Minling. ML-RBF: RBF neural networks for multi-label learning[J]. *Neural processing letters*, 2009, 29(2): 61–74.
- [13] WANG Jiang, YANG Yi, MAO Junhua, et al. CNN-RNN: a unified framework for multi-label image classification[C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA, 2016: 2285–2294.
- [14] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[C]//Proceedings of the 25th International Conference on Neural Information Processing Systems. Lake Tahoe, USA, 2012: 1097–1105.
- [15] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[C]//3rd International Conference on Learning Representations, ICLR 2015. San Diego, USA, 2015: 1409–1422.
- [16] DAI Wenyuan, XUE Guirong, YANG Qiang, et al. Co-clustering based classification for out-of-domain documents[C]//Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Jose, USA, 2007: 210–219.
- [17] ZHUO H H, YANG Qiang. Action-model acquisition for planning via transfer learning[J]. *Artificial intelligence*, 2014, 212: 80–103.
- [18] DENG Jia, DONG Wei, SOCHER R, et al. ImageNet: a

- large-scale hierarchical image database[C]//Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, USA, 2009: 248–255.
- [18] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the Inception Architecture for Computer Vision[C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA, 2016: 2818–2826.
- [19] EVERINGHAM M, VAN GOOL L, WILLIAMS C K I, et al. The pascal visual object classes (VOC) challenge[J]. *International journal of computer vision*, 2010, 88(2): 303–338.
- [20] SHELHAMER E, LONG J, DARRELL T. Fully convolutional networks for semantic segmentation[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2017, 39(4): 640–651.
- [21] ZHANG Chenlin, LUO Jianhao, WEI Xiushen, et al. In defense of fully connected layers in visual representation transfer[C]//Proceedings of the 18th Pacific-Rim Conference on Multimedia on Advances in Multimedia Information Processing. Harbin, China, 2017: 807–817.
- [22] HARZALLAH H, JURIE F, SCHMID C. Combining efficient object localization and image classification[C]//Proceedings of 2009 IEEE International Conference on Computer Vision. Kyoto, Japan, 2009: 237–244.
- [23] PERRONNIN F, SÁNCHEZ J, MENSINK T. Improving the fisher kernel for large-scale image classification[C]//Proceedings of the 11th European Conference on Computer Vision. Heraklion, Greece, 2010: 143–156.
- [24] YANG Jingjing, LI Yuanning, TIAN Yonghong, et al. Group-sensitive multiple kernel learning for object categorization[C]//Proceedings of 2009 IEEE International Conference on Computer Vision. Kyoto, Japan, 2009: 436–443.
- [25] OQUAB M, BOTTOU L, LAPTEV I, et al. Learning and transferring mid-level image representations using convolutional neural networks[C]//Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA, 2014: 1717–1724.
- [26] WEI Y, XIA W, LIN M, et al. HCP: a flexible CNN framework for multi-label image classification[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2016, 38(9): 1901–1907.

作者简介:



余鹰,女,1979年生,副教授,博士,主要研究方向为多标记学习、计算机视觉、粒计算。



王乐为,男,1993年生,硕士研究生,主要研究方向为计算机视觉、深度学习。



吴新念,女,1993年生,硕士研究生,主要研究方向为多标记学习、粒计算。