文章编号：2096-1618(2017)05-0508-05

基于卷积神经网络的课堂表情分析软件研究与实现

周建国， 唐东明， 彭 争， 李 闯， 张加昌
(西南民族大学计算机科学与技术学院，四川 成都 610041)

摘要：课堂上听众的面部表情是听众的心理状态的一个表征，通过分析听众的面部表情数据可以用于评估和改善教学效果。提出利用卷积神经网络分析学生课堂面部表情，进而辅助分析和研究学生们在课堂的专注程度和帮助教师改善教学过程。引入梯度提升方法与卷积神经网络相结合的方式来描述学生表情的图片特征，通过训练神经网络并将其作为图片特征萃取器，使用梯度提升决策树将特征映射至更高维度空间，并融合二者特征进行分类。最后在实际的手机应用中通过使用提出的模型来进行课堂表情数据分析，能够有效地分析课堂上学生的听课情况并帮助教师提升教学效果。

关 键 词：卷积神经网络；梯度提升决策树(GBDT)；特征提取；模型融合
中图分类号：TP391 **文献标志码**：A
doi：10.16836/j.cnki.jcuit.2017.05.008

0 引言

学生在上课过程中情绪的变化可以体现学生课堂上学习的状况，积极的学习情绪可以对大脑主动思维产生强烈的促进作用，从而提高学习能力，而人的情绪又会反应到面部表情上。传统教学过程中，教师通过观察学生的表情然后结合自己的教学经验能够很快知道当前学生的听课状态。现在课堂教学时面对的学生通常较多，教师在授课时很难做到面面俱到观察大多数的学生的情况，并且一堂课时间接近一个小时，教师也难以记住这堂课中学生的课堂听课状态的变化情况。提出利用教师的智能手机来辅助记录课堂上学生的听课情况，并对学生的面部表情进行识别分析，最后综合评估教学过程，提升教学质量。

人脸表情识别系统主要由人脸定位、特征提取、表情分类[1-2] 3大部分组成。传统机器学习方法在面部表情识别方面有所不足，深度学习模型能够有效地弥补不足。在深度学习模型中，卷积神经网络(convolutional neural network, CNN)有强大的图像特征描述能力；可以接受不同尺寸的图像输入，图像经过CNN卷积、下采样、池化操作后，其图片特征能被很好地描述；有了优秀的特征后模型的识别精度会达到很好的水平。人脸表情图像如何分类是人脸表情识别系统的一个重要问题，在对课堂表情进行分类的过程中，尝试使用过ResNet[8]，GoogleNet[9]和AlexNet[10] 3个模型；但因课堂环境复杂造成图片数据质量不高，导致单个模型的分类效果不理想，单个模型并不能很好地描述图片特征；为此尝试细化图片特征，经过研究发现利用梯度迭代决策树(GBDT)将以前的模型学习到的特征映射到更高维度空间后，能在一定程度上补足原有特征的缺陷。然后利用梯度提升方法，将原有特征作为GBDT的输入，对这些特征进行变换，映射到更高维度的空间，从而让模型学习到更多有用的特征。

1 课堂表情分析软件

1.1 基本原理

教学过程中教师通过学生的面部表情来辨别学生是否在专心听课，但是这种方式只能获取到局部范围内学生的上课情况而难以顾全到大部分学生；采用智能手机辅助，不仅能记录下不同时刻学生的上课表情，而且能捕获到课堂中大部分学生的面部神态，方便教师在课后调整教学方案。

通过智能手机采集到课堂上学生上课时不同时刻的面部表情图像后，为了进行分析首先需要提取图像的特征，特征提取直接影响后续表情分析[3-4]。为了更好地描述学生课堂听课时的面部表情特征，采用一种将卷积神经网络(CNN)[5]与迭代决策树(GBDT)[7]相融合的方法提取面部图像特征，通过有监督的方式对采集到的大量训练样本进行人工标签，根据表情将其分为专心与不专心两类样本，然后根据训练样本进行预先训练CNN并将全连接层特征值输入GBDT，接下来训练GBDT并将树的节点作为特征值与CNN特征融合并使用单层隐层MLP感知机进行分类。

梯度提升决策树(GBDT)是一种迭代的决策树算法，该算法由多棵回归树组成，使用梯度下降法求解。

收稿日期：2017-09-06
基金项目：国家自然科学基金资助项目(61003101-131-4)

所有树的结论累加起来作为最终答案。GBDT 的基本思想是通过构建 M 个弱分类器,经过多次迭代最终组合而成一个强分类器。每一次迭代是为了改进上一次结果,减少上一次模型的残差。并且在残差减少的梯度方向上建立新的组合模。提升方法实际采用加法模型(基函数的线性组合)与向前分步算法[11]。以决策树为基函数的提升方法称为提升树,对分类问题的决策树是二叉分类树,对回归问题的决策树是二叉回归树。提升树模型可以表示为决策树的加法模型[11]:

$$f(x)_M = \sum_{m=1}^{M} T(x_j \theta_m) \tag{1}$$

其中 $T(x_j\theta_m)$ 表示决策树, θ_m 表示决策树的参数,M 为树的数目。

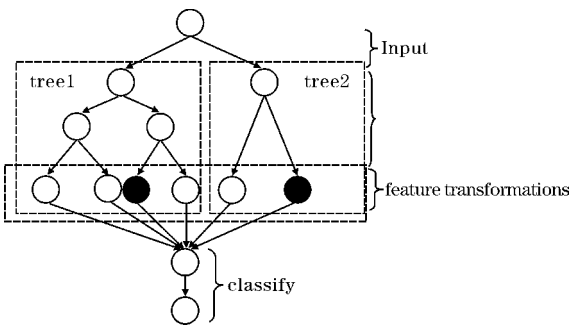

图 1　选择特征的 GBDT 结构图

提升树利用加法模型与前向分步算法实现了学习的优化过程,但对损失函数而言,往往每一步优化并不那么容易,针对这一问题,Freidman 提出了梯度提升算法[6]。接下来用于进行特征选择的 GBDT 结构树如下图 1 所示,从图中可以看出首先采用已有特征训练 GBDT 模型,然后利用 GBDT 模型学习到的树来构造新特征,最后把这些新特征加入预先训练过的模型提取到的特征中,预训练模型采用的是在分类效果相对较好的 AlexNet。最后只使用单隐层的 MLP 进行分类。向量的每个元素对应于 GBDT 模型中树的叶子结点。当一个样本点通过某棵树最终落在这棵树的一个叶子结点上,那么在新特征向量中这个叶子结点对应的元素值为 1,而这棵树的其他叶子结点对应的元素值为 0。新特征向量的长度等于 GBDT 模型里所有树包含的叶子结点数之和。举例说明。图 1 中的两棵树是 GBDT 学习到的,第一棵树有 4 个叶子结点,而第二棵树有 2 个叶子节点。对于一个输入样本点 x ,如果它在第一棵树最后落在其中的第三个叶子结点,而在第二棵树里最后落在其中的第二个叶子结点。那么通过 GBDT 获得的新特征向量为[0, 0, 1, 0, 0, 1],其中向量中的前四位对应第一棵树的 4 个叶子结点,后两位对应第二棵树的 2 个叶子。

1.2　算法流程

根据上述分析,并结合具体的课堂表情识别分析这一问题,提出如下算法:

(1)输入图片 RGB 分量

(2)训练 AlexNet、GoogleNet、ResNet

(3)提取 AlexNet 全连接层特征

(4)AlexNet 特征输入 GBDT 训练

(5)提取 GBDT 特征与 AlexNet 特征相融合

(6)使用 MLP 预测

图 2　手机应用程序运行截图

算法首先利用 Caffe 将每张训练集图片转换成 227×227×3 的像素矩阵,227×227 为图像像素,3 为 RGB 三通道值;并存储为读取效率更高的 LMDB 格式,LMDB 为内存映射数据库,存储格式为 Key-Value;利用 Caffe 将数据通过 caffe. proto 定义的一个 datum 类来封装浮点像素值和通道以及图片标签;lmdb 文件结构简单,一个文件夹,里面一个数据文件,一个锁文件。数据可任意复制、传输且速度极快,利于 I/O 传输;算法第 2 步训练 3 个模型并保存模型结构与学习到的训练参数,避免重复训练模型;保存后的模型只需快速加载即可使用;算法的第 3 ~4 步主要用于提取图片中的表情特征数据,提取特征阶段使用已训练好的 AlexNet 提取图片的纹理特征,每张图片都需要经过 AlexNet 网络;提取采用特征量最大的全连接层每张图片的 4096 维特征,提取后的特征矩阵格式如 $[x_1, x_2, x_3, \cdots, x_{4096}]$,并将其作为 GBDT 的原始特征输入,当数据集在 GBDT 上达到较好的分类效果后再次提取高维度新特征并与原输入特征合并,最后使用 MLP 预测分类。

2　系统实现

系统分为两部分:手机应用程序、后台服务。手机

应用程序提供给教师使用,教师只需在首次使用的时候将学生的基本信息和课程信息导入,进行相应的设置,上课时将手机摆放在合适的位置,程序将会自动拍摄并且不会影响课堂教学秩序。后台数据分析存储服务搭建在阿里云上,并提供 Web 服务。当课堂结束以后,手机应用程序将图像数据提交到后台服务器进行数据分析,教师可以有选择地查看当前或历史的分析结果,通过这些数据教师可以了解每堂课学生听课情况,调整自己的教学策略。除此之外手机应用还具有自动点名功能,历史点名记录也会保存,方便教师随时查看最近的出勤情况。此外应用还有很强的扩展性,例如在采集到的课堂照片中,能获取每个同学的位置信息,可以对每个同学的位置信息进行长期追踪,分析他们的心理情况。图 2 是手机应用程序运行时的部分截图,包括课堂照片采集设置;其中课堂分析结果,横轴为采集的时间点,纵轴为学生课堂专心程度的综合得分;分析历史可以方便教师在日后查阅,给教师提供教学参考。

3 实验及分析

3.1 数据集与主要评价指标

首先为了分析验证模型的有效性,采用公开数据竞赛平台 kaggle 的公开数据集“Kaggle cats and dogs”进行评估。该数据集来自 kaggle 举办的一个竞赛,数据集中有训练数据 25000 张,其中猫狗各占一半,测试集有 12500 张没有标定的图片,该数据集可以在 kaggle 下载。

接下来在真实课堂上采集的数据集上进行分析,该数据来源于教师采用提供的手机应用程序在课堂上进行采集的照片;在数据的采集过程中,进行 17 次课堂照片采集,每堂课采集 30 张照片,每堂课得到的可用人脸数据大约 800 张;最后用于进行实验的数据量为 13600 张;在实验中检测图中人脸,对学生表情进行分析。

因课堂表情数据正负样本比例不均衡,故模型评价指标采用 ROC 曲线,ROC 曲线具有当测试集中的正负样本的分布变化时,曲线能够保持不变的特性。在 ROC 曲线图中横轴为 $FPR=FP/(FP+TN)$,纵轴为 $TNR=TN/(FP+TN)$,AUC(area under curve)表示 ROC 曲线下的面积,其值介于0.1 ~ 1,AUC 数值可以直观的评价分类器的好坏,值越大代表分类器性能越好。为了进行对比,实验将提出的融合特征后的模型命名为 GBM-Net,并与 AlexNet 等模型做对比。在图像分类领域,AlexNet 赢得了 2012 年 Imagenet 图像分类的冠军,掀起了大规模图像识别的开端;而在 AlexNet 后更多更深层的神经网络模型被提出,如 GoogleNet,ResNet 等也在图像识别领域有着同样优秀的表现。

3.2 kaggle 数据集实验结果

在 kaggle 数据集上的 ROC 曲线如图 3 所示,图 3 将 ROC 曲线做了局部放大以辨别各曲线的含义。从图 3可以看出,组合后的模型 GBM-Net 在 kaggle 数据集上分类性能上稍优于其他单个模型。其中,纵轴真正率和横轴负正率在使用不同的分类阈值后得出,GBM-Net 模型的曲线下方面积明显大于其他 3 个模型。

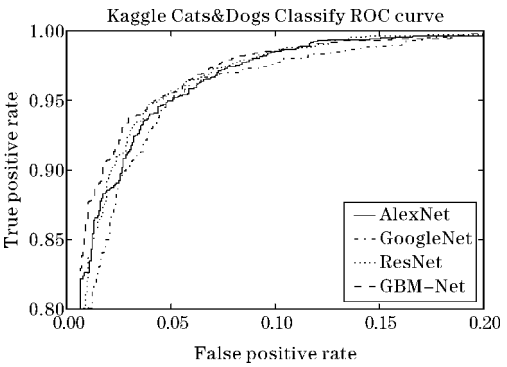

图 3 kaggle 数据集 ROC 曲线

表 1 为 kaggle 数据集上不同模型的参数与得分,所有模型均为在 caffe2 深度学习框架上运行的参数,GBM-Net 经由同一个 AlexNet 提取全连接层特征并通过 GBDT 再次学习特征后的结果,GBM-Net 特征由单隐层 MLP 进行处理并分类。表 1 中 Model 列代表各模型名称;Parameters 组合列中,max_iter 为模型最大迭代次数,综合目标函数的收敛迭代次数与机器性能,所有模型在该阶段均统一迭代次数,base_lr 为初始学习率,视训练数据集大小而定,数据集越大,一般学习率设置越小,gamma 为 base_lr 的衰减系数,达到固定迭代次数时,学习率 base_lr 的值将乘以衰减系数以降低学习率,lr_policy 为学习率下降策略,设置为 step 则影响到 base_lr 的值,momentum 为上一次梯度更新的权重,weight_decay 作为权重衰减项,是防止过拟合的一个参数。在 GBM-Net 模型中,尝试了很多参数的值,在不同数据集上参数取值不同,表 1 中的 hidden_layer_sizes 为隐层神经元个数,activation 为激活函数类型,选择二分类模型选择 logistic 作为激活函数以得到类别概率,solver 为学习方式,采取随机梯度下降(sgd),learning rate 为学习率,设置为 adaptive 自适应;在 GBDT 中,n_estimators 为决策树的数量,其值直接影响到 GBDT 算法的精度,max_features 为每棵子树所能使用最大特征数,设置为 sqrt 则每棵子树只允许使用特征数开方后的值。最后一列 roc-auc score 为模型所得到的分数,其值为图 3 中曲线下方面积。综合图 3 和表 1 的结果可以发现提出的方法获得了最好的结果,验证了模型的有效性。

表 1　kaggle 数据集不同模型算法的对比结果

Model	Parameters						roc-auc score
	max_iter	base_lr	gamma	lr_policy	momentum	weight_decay	
AlexNet	30000	0.001	0.1	step	0.9	0.0002	0.9920
GoogleNet	30000	0.001	0.1	Step	0.9	0.0002	0.9887
ResNet	30000	0.001	0.1	Step	0.9	0.0002	0.9918
	30000	0.001	0.1	Step	0.9	0.0002	
	MLP Parameters						
	hidden_layer_sizes	activation	solver	alpha	learning_rate	max_iter	
GBM-Net	200	logistic	sgd	0.0001	adaptive	2000	0.9920
	GBDT Parameters						
	learning_rate		n_estimators	max_depth			max_features
	0.1		3000	28			sqrt

3.3　真实课堂表情数据集实验结果

在真实课堂表情数据集上的 ROC 曲线如图 4 所示,从图 4 可以看出,组合后的模型 GBM-Net 在真实表情数据集上分类性能优于其他单个模型。从图反映的结果中不难看出,特征融合后,模型的 auc 值明显高于其他 3 个模型;在相对低质量输入图片的情况下,算法能明显提高 AlexNet 的分类性能。同理,表 2 为真实课堂表情数据集上不同模型的参数与得分,不同模型的取值参数如表 2 所示,参数的具体含义和前一个实验表述一致。从表 2 最后一列的 roc-auc score 值结果可以看出提出的模型在表情数据上取得了较好的结果。

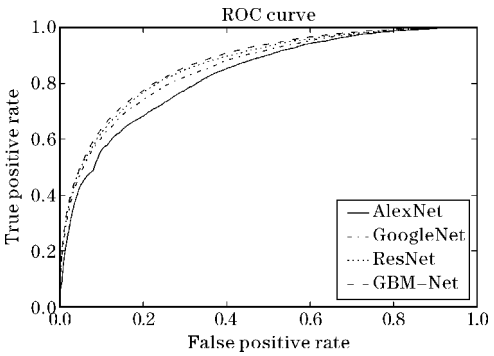

图 4　真实课堂表情数据集 ROC 曲线

表 2　真实课堂表情数据集不同模型算法的对比结果

Model	Parameters						roc-auc score
	max_iter	base_lr	gamma	lr_policy	momentum	weight_decay	
AlexNet	20000	0.01	0.1	Step	0.9	0.005	0.8291
GoogleNet	20000	0.01	0.1	Step	0.9	0.005	0.8545
ResNet	20000	0.01	0.1	Step	0.9	0.005	0.8590
	20000	0.01	0.1	Step	0.9	0.005	
	MLP Parameters						
	hidden_layer_sizes	activation	solver	alpha	learning_rate	max_iter	
GBM-Net	100	logistic	sgd	0.01	adaptive	2000	0.8748
	GBDT Parameters						
	learning_rate		n_estimators	max_depth			max_features
	0.1		2000	15			sqrt

3.4　系统实际使用评估

在教学领域,普遍认同关于教学质量和学习行为的直接和间接评估结果应该一致[12-14]。例如学生在某门课程表现积极专注,那么他很有可能会给这门课程的老师评价较高的分数;相反,如果学生在课堂上表现较为散漫等,则在一定程度上反映了课堂效果不佳。通过该软件的实际运用和相关数据分析可以发现在接近中午下课的时间点,学生听课效果往往不佳;同时学生的课堂注意力大多不会超过半个小时等。因此,通过对软件的分析,还可以进行教师的课程辅助设计:例如教师在备课时候可以将精华重点内容放在课堂的前半部分集中讲解效果更好、在学生注意力较分散时候,多进行和学生的互动交流问答等。

4 结束语

研究了图像识别在课堂辅助教学领域的可行性,探讨了卷积神经网络及特征细化在课堂图像识别方面的作用。为达到更好的分类效果,设计了多模型融合的课堂表情识别模型并提高了模型性能。为将思路运用到实际场景中,同步开发了服务端与客户端;客户端完成了数据采集及分析结果显示等功能,算法模型运行于服务端并将分析结果返回给客户端。目前为止,文中算法仍存在一些不足,在模型准确率及图像质量方面还需要完善,后续将继续研究改进,并尝试加入心理分析,课堂追踪等更多具有实际意义的功能。

致谢:感谢西南民族大学大学生创新项目(S201710656100)、西南民族大学中央高校基本科研业务费专项资金项目(2015NZYQN25)对本文的资助

参考文献:

[1] 张翠平,苏光大.人脸识别技术综述[J].中国图像图形学报:A 辑,2000(11):885-894.

[2] 梁路宏,艾海舟,徐光祐,等.人脸检测研究综述[J].计算机学报,2002,25(5):449-458.

[3] 李华胜,杨桦,袁保宗.人脸识别系统中的特征提取[J].北方交通大学学报,2001,25(2):18-21.

[4] 王聃,贾云伟,林福严.人脸识别系统中的特征提取[J].微计算机信息,2005,21(07X):53-55.

[5] Bouvrie J. Notes on convolutional neural networks [J]. 2006.

[6] Friedman J H. Greedy function approximation: a gradient boosting machine[J]. Annals of statistics, 2001:1189-1232.

[7] Elith J, Leathwick J R, Hastie T. A working guide to boosted regression trees[J]. Journal of Animal Ecology, 2008, 77(4):802-813.

[8] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016:770-778.

[9] Szegedy C, Liu W, Jia Y, et al. Going deeperwith convolutions[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2015:1-9.

[10] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]. Advances in neural information processing systems. 2012:1097-1105.

[11] 李航.统计学习方法[M].北京:清华大学出版社,2012.

[12] 苏耶亚塔·兰尼.用以提升教学效果的情感分析系统[J].计算科学评论,2017,6(1):34-41.

[13] 卢家楣.课堂教学的情感目标分类[J].心理科学,2006,29(6):1291-1295.

[14] Mishra, Brojo Kishore, Sahoo. Abhaya Kumar Source Evaluation of Faculty Performance in Education System Using Classification Technique in Opinion Mining Based on GPU Computational Intelligence in Data Mining. 2016,(2):109-119.

Students′ Expression Analysis in the Classroom based on Gradient Boosting Decision Tree and Convolution Neural Network

ZHOU Jian-guo, TANG Dong-ming, PENG Zheng, LI Chuang, ZHANG Jia-chang

(1. School of Computer Science and Technology, Southwest Minzu University, Chengdu 610041, China)

Abstract: The facial expression of the students in the classroom is a representation of the mental state of the students, and the facial expression data of the audience can be used to assess the teaching effect. This paper proposes the use of CNN(convolutional neural network) to analyze students`facial expressions, and then study students facial expression so as to helping teachers improving the teaching process. In this paper, we combine GBDT(gradient boost decision tree) and CNN to describe the characteristics of the pictures. By training the neural network and using it as the image feature extractor, we use the GBDT to map the feature to higher dimension space and then according to the characteristics of the two categories to classified. Finally, in the practical application, it can effectively analyze the lectures of the students in the classroom and help the teachers to improve the teaching effect.

Keywords: convolution neural network; gradient boosting decision tree; feature extraction; model fusion