

Behavior analysis using Resnet

Peizhi Wang

2018年3月6日

摘要

This is simply a recording of my experiments on students behavior analysis. First week i will use Resnet to train all my training data. Then i will consider using transfer learning, data augmentation and other methods to build my model.

1 Introduction

教育永远是一项非常重要的事业，是人类文明继承发展的主要方式。当今设备教育行业存在着如下几个问题：

1 基础教育过分的注重考试成绩，给学生们增加了很大的压力

2 大学教学只要考试合格就行，教育质量大大下滑

为了解决上述两个问题，一个非常可行的方向是记录学生的平时成绩和课堂活跃程度，这样既可以缓解基础教育同学们的考试压力，同时可以提高大学教育的质量。

但是，通过老师和助教来记录平时成绩是非常耗费精力的，而且由于部分助教不认识某些学生，就很难来记录学生的表现。如果点名的方式就会浪费珍贵的课堂时间。

然后，大学教室和大部分高中课堂是有着摄像头的。而今年的机器视觉又赋予了计算机来理解图片的能力。所以本文提出一种利用深度残差网络来对学生行为进行分类的方法，在复旦大学采集了一定量的数据集来进行训练和测试，取得了

还可以追踪学生的情绪，并且在学生情绪不佳的时候采取干预。

另外，通过对具有一定比例的学生的行为的分析，也可以用来评价老师的上课质量和学生对老师上课的满意程度。通过评教的方法是具有很大的误

差，有时候学生也会因为被公开的评价会引起老师的生气，所以学生不容易在问卷调查中给出客观的评价。

2 Patent using grading

基于人脸和语音识别的课堂欣慰监控系统及方法自动识别不同的教学场景自动识别不同的上下课时间

1.一种基于人脸和语音识别的课堂行为监控系统,其特征是,包括

视频信息采集系统,包括安装在教室四个墙角的全方位旋转摄像头、与摄像头相连的解码器、与解码器相连的画面分割器,用于采集课堂中学生、教师的视频数据信息;

语音信息采集系统,包括安装在学生课桌下面以及讲台上的录音设备,用于采集课堂中学生、教师的语音数据信息;

主控处理器,对学生、教师的视频数据信息和语音信息进行预处理,提取学生、教师的面部表情特征和行为特征;

分析处理器,将学生、教师的面部表情特征和行为特征于标准模板进行比较,并计算相应得分。

2. 一种基于人脸和语音识别的课堂行为监控方法,其特征是,包括以下步骤:

步骤一、通过安装在教室四个墙角的摄像头采集课堂中学生、教师的视频数据信息;通过安装在学生课桌下面以及讲台上的录音设备采集课堂中学生、教师的语音数据信息;

步骤二、采集到的学生、教师的视频数据信息经解码器解码和画面分割器分割后,发送给主控处理器;将采集到的学生、教师的语音数据信息发送给主控处理器;

步骤三、主控处理器对接收到的学生的视频数据信息进行预处理,提取学生的面部表情特征和行为特征,通过分析处理器将学生面部表情特征与学生面部表情标准模板进行比较,依据比较结果计算课堂中学生面部表情表现得分,将学生行为特征与学生行为标准模板进行比较,依据比较结果计算课堂中学生行为表现得分;

步骤四、主控处理器对接收到的教师的视频数据信息进行预处理,提取教师的面部表情特征和行为特征,通过分析处理器将教师面部表情特征与及教师面部表情标准模板进行比较,依据比较结果计算教师对学生课堂表现面部表情反应的得分,将教师行为特征与教师行为标准模板进行比较,依据比较结果计算教师对学生课堂表现行为反应的得分;

步骤五、主控处理器对接收到的学生的语音数据信息进行处理,提取学生语音特征,训练语音标准模板,将学生语音特征与语音标准模板进行比较,依据比较结果计算课堂中每位学生的发言次数及频率、发言时间长短以及小组讨论时的发言比率;

步骤六、主控处理器对接收到的教师的语音数据信息进行处理,提取教师语音特征,计算教师教学效果的分值,并与教师教学效果平均分进行比较,当该分值小于教师教学效果平均分,发出提示;

步骤七、主控处理器将课堂中学生面部表情和行为表现得分以及教师对学生课堂表现面部表情反应和行为反应的得分进行整合,得出每个学生的课堂行为总分,并将该总分与主控制器中设置的学生课堂行为平均分数进行比较,当该总分小于学生课堂行为平均分数,发出提示;

步骤八:将学生、教师的视频数据信息、语音数据信息、每个学生的课堂行为总分和教师教学效果的分值存储到主控处理器的数据库中。

5. 根据权利要求2所述的一种基于人脸和语音识别的课堂行为监控方法,其特征是,所述步骤三中,在进行学生的面部表情特征和行为特征与标准模板进行比较之前,先对学生面部表情标准模板和学生行为标准模板进行赋分,所述学生面部表情标准模板中专注、高兴表情设定为10分,冷漠表情设定为4分,烦躁表情设定为1分;学生行为标准模板中举手、低头做笔记、抬头听课的行为设定为10分。

6. 根据权利要求2所述的一种基于人脸和语音识别的课堂行为监控方法,其特征是,所述步骤四中,所述教师面部表情标准模板包括嘴角弧度、眉内点和眉外点;所述教师行为标准模板包括教师点头次数。

7. 根据权利要求2所述的一种基于人脸和语音识别的课堂行为监控方法,其特征是,所述步骤四中,在进行教师的面部表情特征和行为特征与标准模板进行比较之前,先对教师面部表情标准模板和教师行为标准模板进行赋分,所述教师面部表情标准模板中愉悦表情设定为10分,不满表情设定为0分,其余在中间范围依据满意程度赋分,所述教师行为标准模板中依据教师点头次数进行赋分。

8. 根据权利要求2所述的一种基于人脸和语音识别的课堂行为监控方法,其特征是,所述步骤五中,学生语音特征包括学生每堂课的发言次数和频率、每次发言时间长短和小组谈论时发言比率。

9. 根据权利要求2所述的一种基于人脸和语音识别的课堂行为监控方法,其特征是,所述步骤五中的训练语音标准模板具体方法为:根据每个说话人的训练语音样本,经特征提取,建立每个说话人的语音标准模板。

10. 根据权利要求2所述的一种基于人脸和语音识别的课堂行为监控方法,其特征是,所述步骤六中,计算教师教学效果的分值的具体方法为:按照语音标准模板将教师的语音信息分割为若干个小单位,利用多元化测度法计算得到其测度值,根据分析处理器中标准语音库建立语音评分模型,将获得的测度值转换成为评价教师教学效果的分值。

背景技术

[0002] 对于课堂行为监控,是学校评价教学质量的重要环节,充分了解教师的教课水平和学生上课的反应,才能保证高质量教学水平。现有,对于课堂行为的监控采用学生记录或者老师测试模拟、老师观察监督的方式,此种方式不能够充分调动学生的学习兴趣,不能评价老师的教学效果,无法同时实现对学生、教师的课堂行为的采集、分析、记录及评价。因此,如何利用基于人脸识别技术和语音识别技术来对学生和教师的课堂行为进行精确采样和智能分析与评价,实现对教学课堂中学生和教师的表现行为进行观察和记录,有效提高课堂教学效果,促进学生的全面发展是值得研究和开发的课题。

3 Patent for teacher

1将学生的评价数据采集起来,通过分析学生的行为,通过长期的累加,对教师的课堂教学质量进行评估。

在传统的教育活动中，人们往往过分强调老师教的作用，而忽略了学生听的效果，这种割裂教学的模式，往往造成学生积极性地下，缺乏上课的即使激励。然而现今教师的授课方式往往只有一名老师，在规定时间内对多名学生进行某一领域知识的教授。教师在有限的时间与精力下，并不能在课堂上完全了解每个学生对授课内容和质量的反应，为此，教师往往在课后对学生进行教学质量的调查。

[0002] 近年来，在教育模式不断扩大的过程中，高等教育的教学质量日趋重要，为保证教学达到预期的目的，需对教学的全过程进行检查、评价、反馈、控制和调节。然而，据调查结果统计，很多高校的课堂教学质量还停留在手工统计数据、手工计算分析监控课堂教学的层次上，需要投入大量的人力物力，且学生家长没有及时得到学生每次课的出勤情况，不能及时监督学生。

[0003] 上述这些缺点的存在，表明传统的教学监控系统已经不能有效的监控教学，有必要研制出一种能够更有效、全面反映教学过程中出现的问题，以便教学监管部门监控和调节教学，和及时反馈给学生家长每次课学生的出勤情况的教学监控系统。

[0004] 目前，课堂教学质量调查主要是通过调查问卷的方式，即在授课结束后，教师向学生分发课堂教学质量调查问卷；学生根据个人对本节课的感受与评价，填写调查问卷，也可自主提问问题，或记录对课堂教学的建议或意见等；学生在填写完调查问卷后，教师将调查问卷进行回收，然后，利用课下时间对调查问卷的内容进行统计与分析。

[0005] 但是，这种调查方式存在多个不便捷之处，第一，打印调查问卷耗时、耗材；第二，统计调查数据的过程将耗费教师大量的时间与精力；第三，人为统计调查数据的过程中，不可避免的将出现一些人为差错；第四，每次的评分结果不易保存，不利于长期的对某教师教学质量的统计分析。

[0003] 用计算机和网络技术实现的基于在线问卷的教师评价系统，方便了学生评价和对评价的存储、计算、统计、分析、显示等自动化处理。传统的基于文字描述的教师评价方法，由于评价数量少、内容随意性大，很少运用文本分析技术进行处理。基于在线问卷的方法还在市场调研、满意度调查等领域获得广泛应用。

[0004] 但基于问卷调查评价教师的方法在有效性方面具有局限性：(1) 一次性，每个学生只对教师的整个学期或课程情况在学期或课程结束后评论一次；(2) 数据量小，每个学生只对一个教师评论一次；(3) 限制性，调查问卷方法的客观性强、时间固定，不便表示特定时间的真实情绪，不能自由表达对教师的真实感受。

出于对机器视觉准确率率的考量，所以就采用开心愉悦10分，冷漠4分，焦躁愤怒0分的打分制度。

发明内容

[0005] 本发明的目的在于针对现有技术的缺陷和不足,提供一种结构简单,设计合理、使用方便的基于情感分析的教师评价系统,针对现代学生评论教师的表述特点,以8个基本情感为基础,快速、动态地建立情感数据字典SDD,对异构多源、含有情感符号的大量的短文本,把情感极性与情感强度的计算融合在一起进行情感分析,并通过基本情感的合成,多方面、多方位对教师进行综合性的量化评价。

[0006] 为实现上述目的,本发明采用的技术方案是:

[0007] 一种基于情感分析的教师评价系统,包含情感数据词典(SDD)、SDD建立与维护模块、信息采集及预处理模块、情感分析模块和教学评价模块;

[0008] 所述情感数据词典(SDD):基于Robert Plutchnik[1]的情感理论,以8个基本情感及其词语——信任、期待、愉悦、惊奇、伤心、厌恶、恐惧、愤怒,为基础;这8个基本情感分别表现了相互对立、相互矛盾的正面和负面的、积极和消极的基本观点或倾向,他们表示出不同的情感倾向和情感强度;在一个评论中他们的不同组合会有不同的情感倾向和情感强度;不同组合情感的组合还会产生不同的情感倾向和情感强度;本发明把情感极性与情感强度的计算融合在一起,通过迭代计算,同时得到情感倾向和情感强度;

收集的信息包括文字,视频,音频,图像。这里考虑先进行图像的研究,然后拓展到视频方面。图片采样可以每0.1s采样一次,这样就可以有10FPS的照片数量。

获取班级位置的分布,比如是正三角,倒三角,均匀分布,前排空后排集中式等。通过分析学生上课历史的课堂特点,以及任课老师历史上课特点,定性来判断学生和老师的特点和行为。

[0022] 座位位置信息;

[0023] 获取班级位置的分布,比如是正三角,倒三角,均匀分布,前排空后排集中式等,通过分析学生上课历史的课堂特点,以及任课教师历史上课特点,定性判断学生与教师的特点和行为;获取学生的位置信息,综合分析学生的个人座位喜好,通过对学生历史上该课程的座位情况,间接反映学生的学习习惯、生活习惯;获取学生附近人员位置信息,比如和谁同桌,以及最相邻学生信息,通过多次数据统计反映学生的人际交往情况;

学生的表情,微笑或者严肃等,间接反映老师课堂中的兴奋点,课堂活跃程度。

检测学生是否处于睡眠状态,将信息反馈给教师,以便教师即使调整教学方法,教学内容以达到更好的教学效果,提高远程教学环境的交互性和教学质量。

智能教师现代远程教育的若干困难。

信息社会,只是爆炸,人们需要终身学习。这是个最好的时代,以网络为载体的远程教育顺应了这一趋势,给人们随时随地获取新知识提供了便利

和强有力的支持。

但是这个方式也存在一些问题，比如学习是学生自发进行的，学生很难受到监督，这对于可以给出课程证书的远程课程就会减少其有效性。而且由于缺乏学生给与老师的即时反馈，老师也很难实时的发现学生在学习上存在的问题。

1.2.4 检测方案比较

远程学生和现场学生在智能教室上课时,可能由于疲劳等原因而注意力不集中,甚至出现瞌睡现象,主要有以下几种情形:

- a). 学生身体直立, 闭眼睡觉;
- b). 学生身体直立, 不断点头瞌睡;
- c). 学生身体伏在课桌上睡觉;
- d). 学生手放在课桌上支撑着头睡觉;
- e). 学生身体倚靠在座椅上, 头部后倾睡觉。

针对以上场景可采集利用的信息有:

- a). 眼睛状态:通过眼睛睁开闭合状态的检测确定学生是否已经处于瞌睡状态;

- 7 -

- b). 头部姿势:点头的动作和瞌睡的相关系数仍然没有找到合适的关系, 监测头部运动的准确率不高;
 - c). 压力:桌面上物品的堆积也可产生对桌面的压力, 难以区分;
 - d). 遮挡:类似压力信息, 多种情况都可造成遮挡, 区分度不高, 难以判定是否是学生伏于桌面瞌睡引起的遮挡;
-

4 Important multi hot code and split CNN

想法的出发点是学生的动作肯定不是独立的.也就是说可以一边听课一边写作业,那么使用One-Hot编码的学生行为数目就比较小。

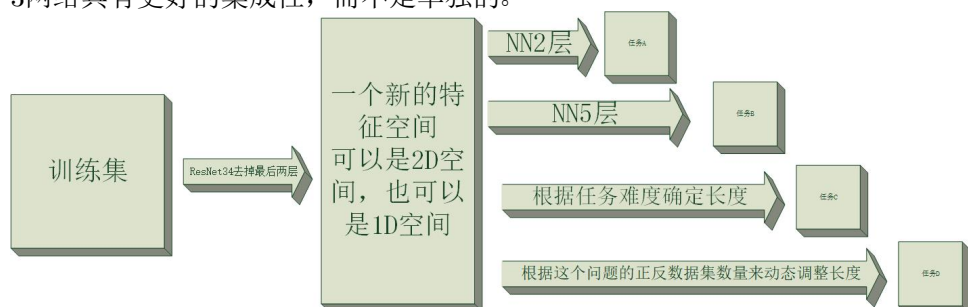
所以这里就放开One-hot的限制,使用Non maximun编码,也就是说让网络挑选出学生最有可能的至多三种行为,对于后面的行为我们就认为没有了。一种proposal是采用多个神经网络对不同行为来进行分类。比如A网络来判决是否认真听课, B网络来判决是否有在吃东西, C网络判决是否睡觉。

后来我想了一种方法就是采用分叉CNN的方法,也就是一开始所有数据先经过一些网络层,提取出一部分特征,分布在M-Domain中,然后每个行为各自分叉一支出去。这样的好处有:

1网络的计算量变小了,共享了一部分的FP计算,视屏监控对实时性要求比较高

2网络具有更好的泛用性,借鉴了迁移学习的想法,形成一种共享学习的概念,可以减少对训练数据集合的要求。不然会随着任务数指数爆炸。

3网络具有更好的集成性,而不是单独的。



5 Ideas

可以加入学生的远景和近景,分辨学生的位置

可以加入学生的表情: happy, sad, 没睡醒

可以加入学生的性别和身份识别

区分上下课: 上课做的整整齐齐, 下课都是随意走动

可解释性: 统计大部分时间是0输出的神经元的个数, 统计这个和Variance/Bias tradeoff的关系

我觉得可以尝试的一个方向是给教室里的每个学生做逐像素点的标记

6 ZNN:Zoom Nerual Network

我在想根据训练和测试图像，放大缩小结构大小或者卷积核大小
或者我在想能不能把卷积核弄成圆形的???

Inception 可以去看看

7 Action Type

根据某个观察量表，分为三大维度

一个是投入状态，包括一般倾听，举手发言等

另一个是非投入状态，包括看课外书，吃东西，玩手机，做作业，趴桌子，
低头，使用PC电脑，喝水，左右交头接耳

另外还有课堂异常行为，包括缺席，中途站起离场或擅自走动，

8 Action Recognition V1.0

today i try to implement my first prototype V1.0.It is based on ResNet
by(he kaiming).The training data is about 180 pictures about myself doing
four kinds of actions.

without data augmentation,after 86 epochs,the training process was early
stoped. we get nearly 1.00 accuracy on training data and 0.90 accuracy on
validation set,which is quite satisfying.

数据增强=False

使用了ResNet18

$$Rows = 612$$

$$Cols = 612$$

$$EarlySTOP = 87$$

$$acc_t = 1$$

$$acc_v = 0.9$$

主要是四个动作，睡觉写字玩手机听课

9 Action Recognition V1.1

数据增强=True 主要是每个图片的零均值单位方差，然后加上ZCA白化
使用了ResNet18

$$val/all = 0.15$$

$$Rows = 612$$

$$Cols = 612$$

$$acc_t =$$

$$acc_v =$$

$$acc_t/sum = 0.5$$

当我开启ZCA的时候，遇到了MEMORY ERROR，可能是在计算zca的时候会生成大矩阵的缘故吧

突然想到，也许可以比较一下远近的识别，也可以把近处的同学和远处的同学都区分开来，证明这个系统的泛用性。

第一次实验由于不小心early stop了，效果并不好。

10 Action Recognition V1.2

使用了ResNet34，观察网络层数增加的影响

11 Action Recognition V1.3

使用了新的数据集dataset2

A1 - A60 一边玩手机一边听音乐,低头

B1 - B58 一边看书一边听歌，低头

C1 - C55 站立，抬头

D1 - D62 举手，抬头

E1 - E60 抬头，喝水

F1 - F71 抬头，吃蛋糕

G1 - G60 抬头，写字

H1 - H55 一边看书一边写字

J1 - J67 一边玩手机一边看书
K1 - K52 吃东西，看书，低头
L1 - L70 低头，看书
M1 - M61 拿起手机在看
N1 - N66 趴着睡觉
O1 - O28 用手撑着睡觉
P1 - P37 抬头听课

12 Residual Network

$$y_l = h(x_l) + F(X_l, W_l)$$

$$X_{l+1} = f(y_l)$$