

一种基于姿势识别的教师教学行为评价方法

郑誉煌

(广东第二师范学院教务处, 广东 广州 510303)

✉zhyhaa@126.com



摘要: 随着智慧教室的不断普及, 采用传统人工评教方法已经不能满足评估教师在这些教室教学视频中教学行为的需要。针对这些教学视频, 建立了智能化评价教师教学行为的框架。这个框架首先基于HRNet深度学习网络获得教师人体姿态信息, 然后根据教师的姿态信息建立了评价其教学行为的指标, 最后采用模糊综合评价的方法实现了对教师教学行为的综合评分。实验表明, 本框架的评价结果总体和传统的人工评价结果是一致的。

关键词: 教师教学行为; 深度学习; 模糊综合评价; HRNet模型; 姿势识别

中图分类号: TP391.41 **文献标识码:** A

An Evaluation Method for Teaching Behaviors based on Posture Recognition Algorithm

ZHENG Yuhuang

(Academic Affairs Office, Guangdong University of Education, Guangzhou 510303, China)

✉zhyhaa@126.com

Abstract: With the continuous popularity of smart classrooms, traditional manual evaluation methods are far from meeting the needs of evaluating teaching behaviors in teaching videos of smart classrooms. In view of this problem, this paper proposes a framework for intelligently evaluating teaching behaviors. In the framework, teachers' body posture information is first obtained through HRNet (High Resolution Net) deep learning network. Then an index to evaluate their teaching behavior is established according to the teachers' posture information. Finally, comprehensive scoring of the teachers' teaching behavior is achieved by using the fuzzy comprehensive evaluation method. Experiments show that the overall evaluation results of the proposed framework are consistent with traditional manual evaluation results.

Keywords: teaching behaviors; deep learning; fuzzy comprehensive evaluation; HRNet model; posture recognition

1 引言(Introduction)

教师教学行为评价是指学校对教师在课堂上的教学行为有效性进行评价的过程。将评价结果反馈给相关教师, 可以指导他们课堂教学能力的提升。当前, 教师教学行为评价在很大程度上依赖于学校管理人员、教学督导或其他专业教师的课堂观察。这种评价方法只能对教师具体的一节课堂教学进行评价, 却难以对教师在整个学期的全部课堂教学进行观察, 不能对教师的课堂教学形成客观和长期的分析。研究新

的教师评价方法和框架成为推动教学改革的新动力之一。当前, 人工智能在教育领域的应用成为教育技术领域的新兴研究课题之一。将人工智能技术应用于教师教学行为评价, 更是一个新兴的研究领域。随着智能教室等教育信息基础设施不断投入使用, 实时和全过程地获取教师教学行为成为可能。利用人工智能技术实现对教师教学行为的全过程智能化分析, 为教师提供客观的课堂评价数据, 对促进教师课堂高效教学和提高课堂教学质量具有重要的意义^[1]。

2 相关研究(Related works)

2.1 教师教学行为的视频分析与评估

视频分析法是采用智能化的手段跟踪视频的目标对象,并获得目标数据的运动数据。HUANG等^[2]提出了一种基于深度卷积神经网络和级联的人脸特征点定位方法,分析和识别学生头部姿态和面部表情,获取学生的课堂行为,从而评估教师上课效果。闫晓炜等^[3]基于yolo算法完成了教师上课时两种肢体动作的识别,周鹏霄等^[4]基于OpenCV的帧差面积法分析教师上课时提问和板书两种行为。视频运动目标分析法采用非接触式的运动分析法,对被检测对象没有产生额外干扰,非常适合应用在课堂教学环境中,在运动教学中得到一定的运用,然而当前教师课堂教学中能识别的动作比较少。

获得教师教学行为数据后,需要评估这些行为反映的教师教学状态和教学效果,不少最新研究都集中在建立各种教学评估模型上。BLACKWELL^[5]采取专家组和非专家组观看及评价乐器教学演奏视频的方法,研究发现专家组和非专家组的评价结果总体是匹配的,而且不随教学经验的变化而变化。GREVE等^[6]提出了一个用于分析课堂视频的评分系统,然而评分来自被授权的观察员对教师的主观评估。上述研究存在无法提供定量分析结果的问题,而且有些定量数据的来源也是通过其他人员的主观评估而得。教师教学行为的评估还需进一步和教师教学行为数据自动对接,才能发挥信息技术的巨大优势,实现教师教学行为分析与评估全程自动化。

2.2 项目的研究框架

人体姿势反映了人体运动特征、人的情绪等综合信息。在教学中,教师的肢体语言充分反映了教师的教学状态,对学生学习兴趣具有重大的影响。本研究建立了基于深度学习技术的课堂教学行为的分析和评估框架,这个框架采用HRNet深度学习网络建立教师课堂教学行为的人体姿态估计模型,从而获取教师上课时的人体关键点;分析这些人体关键点与教师课堂教学行为的对应关系,建立评价教师教学行为的指标;采用模糊综合评价法对这些指标进行综合评估,从而判断教师教学效果。本研究成果能做到教师全过程教学评估,提高评价效率,对有效规范教师教学行为、促进教师教学能力提升,从而提升教学质量具有一定的帮助。

本项目的14个测试视频来源于视频网站haokan.baidu.com公开的教学示范课,每个示范课时长为8—12分钟,视频展示了教师在黑板前的教学活动情况。这些视频覆盖了小学和初中的语文、数学、英语等主干课程,部分视频还展现了一些教师容易出现的负面教学行为。文献[5]的研究方法,本项目组还邀请了三位资深教师对每个视频独立打分,评估视频中教师教学效果,重点考查教师教学行为。对比教师们的评估结果与本文算法的分析结果,从而验证本算法的有效性。

实验平台的硬件环境: Intel i7-8750H@2.20 GHz CPU, 32 GB内存, Nvidia RTX 2070 GPU, 8 GB显存;

软件环境: CUDA Toolkit V10.0, CUDNN V7.0, 基于Anaconda 3的Python V3.7, PyTorch 1.6深度学习框架, Windows 10 64 bit操作系统。HRNet模型权重采用精度最高的官方预训练权重COCO w48 384×288, HRNet模型计算通过GPU加速。

3 人体姿势识别算法(Human posture recognition algorithm)

3.1 项目的研究框架

人体姿态估计被定义为图像或视频中人体关节(也被称为关键点,例如肘部、手腕等)的定位问题。人体骨骼关键点的识别在人体姿势识别的应用中有着举足轻重的理论价值,它是该任务的基础方法之一^[7]。姿势识别主要基于骨骼关键点检测来实现,因为骨骼关键点的检测更能精确地描述人体姿势。基于深度学习的人体姿态估计核心在于提高人体骨骼关键点检测精确度。在当前众多人体姿势识别模型中,HRNet模型的精度是最高的^[8]。

HRNet模型是中国科学技术大学和微软亚洲研究院共同研究并发布的人体姿态估计模型,是CVPR 2019(Conference on Computer Vision and Pattern Recognition, 国际计算机视觉与模式识别会议)影响力最大的10篇论文之一。HRNet模型采用与众不同的并联结构,可以随时保持高分辨率表征,保证姿势识别的效果稳定输出,在COCO公开数据集的关键点检测、姿态估计、多人姿态估计这三项任务里,HRNet模型都超越了现有的人体姿态估计模型。因此本文采用HRNet模型作为人体姿态估计算法核心,算法原理如图1所示。

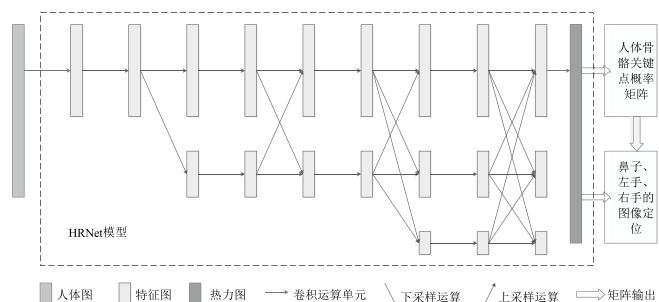


图1 人体姿态估计算法原理

Fig.1 Principle of human posture estimation algorithm

HRNet模型结构如图1虚线部分所示,横向表示模型深度变化,纵向表示特征图尺度变化。第一、第二、第三行分别负责高、中、低分辨率特征图的多层卷积运算,每行在适当的卷积运算单位之间分别进行上、下采样运算,以此相互交换信息,实现多尺度融合与特征提取。最终所估计的人体骨骼关键点是高分辨率的卷积运算,形成输入人体图的骨骼关键点热力图。根据热力图,可以获得鼻子、左右眼、左右耳、左右肩、左右肘、左右手、左右臀、左右膝、左右脚,共17个骨骼关键点的位置信息和最优估计概率。教师的手势和脸部对教学效果影响最大,因此本研究着重分析鼻子和左右手这三个骨骼关键点,如图2中黑色方块所示。

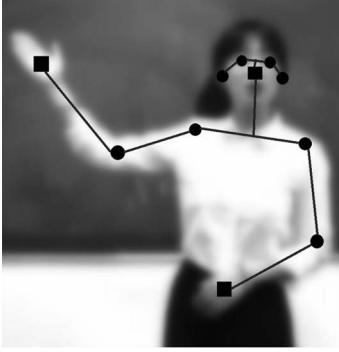


图2 测试视频中的人体骨骼关键点示例

Fig.2 Example of human bone points in test video

给定一段教学视频,可定义为由帧率为 fps 的 n 幅图片组成的集合 $\{image_1, \dots, image_n\}$ 。把HRNet模型作为一个算子 $HRNet(\cdot)$,第 i 帧图片中的鼻子(nose)、左手(left_hand)和右手(right_hand)估计概率 p 和位置信息 (x, y) ,如公式(1)所示:

$$\begin{bmatrix} p_{nose,i} & x_{nose,i} & y_{nose,i} \\ p_{left_hand,i} & x_{left_hand,i} & y_{left_hand,i} \\ p_{right_hand,i} & x_{right_hand,i} & y_{right_hand,i} \end{bmatrix} = HRNet(image_i), i = 1, \dots, n \quad (1)$$

3.2 有效关键点的选择

教师在教学过程中,肢体动作不停地变化,姿势不停地变换,例如转身、遮挡、肢体交互等,导致鼻子和左右手可能在图像中暂时消失,这三个关键点的对应估计概率比较低。设判断鼻子消失的关键点概率阈值为 p_{th_nose} ,则全部满足 $p_{th_nose,i} > p_{th_nose}$ 的鼻子位置数据组成一个有效显示鼻子关键点的定位序列 $Q_{nose} = \{(x_{nose,i}, y_{nose,i}), (x_{nose,j}, y_{nose,j}), \dots\}$ 。

同理,设判断左右手消失的关键点概率阈值分别为 $p_{th_left_hand}$ 、 $p_{th_right_hand}$,可得有效显示左右手关键点的定位序列 Q_{left_hand} 、 Q_{right_hand} 。一般地,取 p_{th_nose} 、 $p_{th_left_hand}$ 、 $p_{th_right_hand}$ 均为0.7,以保证关键点能被正确识别。图3是测试视频1的鼻子和左右手定位信息展示。

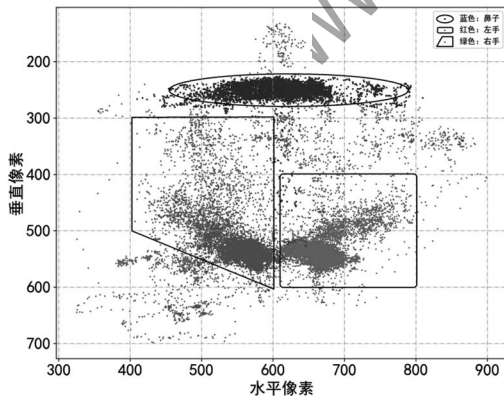


图3 测试视频1的鼻子和左右手定位信息展示

Fig.3 Display of nose, left & right hands positioning information in 1st test video

4 教学行为状态的评价指标(Evaluation criteria of teaching behaviors)

根据上节所获得的测试视频鼻子、左右手定位序列 Q_{nose} 、

Q_{left_hand} 、 Q_{right_hand} ,进一步定义教师教学行为状态的评价指标。

4.1 教师板书的评价指标

对于一个测试视频序列 Q_{nose} ,若相邻的两个元素的帧数差大于等于两倍的帧率,即满足公式(2)条件,则记为教师完成一次有效板书:

$$Q_{nose,k+1} - Q_{nose,k} = j - i \geq 2fps \quad (2)$$

按此定义,教师一次连续板书时间要超过2秒或以上才能算是完成一次有效板书。这次的有效板书时长如公式(3)所示:

$$t = \frac{j-i}{fps} \quad (3)$$

设这个测试视频的有效板书时长序列为 $\{t_1, t_2, \dots, t_m\}$,可得评价指标1—3如下:

指标1:有效板书次数 $C = \text{card}\{t_1, t_2, \dots, t_m\} = m$ 。

指标2:最长有效板书时长 $t_{\max} = \max\{t_1, t_2, \dots, t_m\}$ 。

指标3:平均有效板书时长如公式(4)所示。

$$t_{\text{mean}} = \frac{1}{m} \sum_{i=1}^m t_i \quad (4)$$

4.2 教师头部与手部的评价指标

对于一个测试视频序列 Q_{nose} ,设鼻子的平均水平位置是 $x_{nose,mean}$,平均垂直位置是 $y_{nose,mean}$,则头部偏向左边时间是 $t|_{x_{nose} > x_{nose,mean}}$,头部偏向右边时间是 $t|_{x_{nose} < x_{nose,mean}}$,可得指标4。

指标4:头部左右偏差时间如公式(5)所示。

$$\Delta t_{nose} = \left| t|_{x_{nose} < x_{nose,mean}} - t|_{x_{nose} > x_{nose,mean}} \right| \quad (5)$$

教师在教学中手部超过胸口以上位置的手势,称为上区手势^[9]。手势在这一区域活动,一般表示教师基本进入教学状态,而且情绪比较激昂。设左右手的平均垂直摆幅为 $y_{hand,mean}$,则上区手势阈值定义如公式(6)所示。

$$y_{thr_hand} = (y_{hand,mean} - y_{nose,mean}) \times 0.3 + y_{nose,mean} \quad (6)$$

指标5:左手属于上区手势的帧数 $frame_{left} = \text{card}\{y|y_{left_hand,i} < y_{thr_hand}\}$ 。

指标6:右手属于上区手势的帧数 $frame_{right} = \text{card}\{y|y_{right_hand,i} < y_{thr_hand}\}$ 。

5 教学行为综合评价(Comprehensive evaluation of teaching behaviors)

运用上述教师教学行为评价指标,对本文的14个教师教学视频进行综合评价,以评估视频中教师的教学行为。本研究采用模糊综合评价法^[10],此方法以模糊数学为理论基础,它既可以把对象的多个评价指标融合分析获得其综合分值,还可以把综合分值按最大隶属度原则去评定对象的等级,尤其适于教师教学的评价。

模糊综合评价法包括确定因素集、权重集,计算因素评判矩阵,综合评价计算。

本项目中,因素集即指标1—6,其对应权重集是 $\{0.2, 0.2, 0.2, 0.2, 0.1, 0.1\}$ 。权重集是因素集中每个指标在

综合评价里面所占的权重。由于教师板书时包含了一部分上区手势，因此左右手的上区手势权重应适当降低。

设某一个评价指标在14个测试视频中的平均值是 g_{mean} ，最大值是 g_{max} ，最小值是 g_{min} ，则取平均值 g_{mean} 作为“最优秀”，建立公式(7)的三角形隶属度函数。这个隶属度函数体现了该指标越接近平均评价水平，教师教学行为就越好。根据公式(7)，计算14个测试视频中指标1—6的隶属度，形成因素评判矩阵。

$$h(g)=\begin{cases} \frac{g-g_{min}}{g_{mean}-g_{min}}, & g_{min} \leq g \leq g_{mean} \\ \frac{g_{max}-g}{g_{max}-g_{mean}}, & g_{mean} \leq g \leq g_{max} \\ 0, & \text{其他} \end{cases} \quad (7)$$

因素评判矩阵乘以权重集即可得到最后的综合评价，如表1所示。计算综合评价与三位专家评分均值之间的相关系数，可得相关系数是0.99，即两者的相关程度非常大，表明本算法对教师教学行为的评价和专家的评价基本一致。可见，本算法是有效的。

表1 因素评判矩阵与综合评价

Tab.1 Factors evaluation matrix and comprehensive evaluation

视频序号	因素评判矩阵						综合评价	三位专家评分均值
	C	t_{max}	t_{mean}	Δt_{nose}	$frame_{left}$	$frame_{right}$		
1	0.40	0.57	0.53	0.09	0.32	0.71	0.42	0.47
2	0.92	0.51	0.54	0.29	0.00	0.70	0.52	0.53
3	0.79	0.75	0.68	0.62	0.48	0.71	0.69	0.72
4	0.81	0.74	0.91	0.00	0.00	0.99	0.59	0.62
5	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
6	0.66	0.26	0.37	0.53	0.00	0.72	0.44	0.46
7	0.92	0.41	0.87	0.91	0.00	0.75	0.70	0.69
8	0.13	0.70	0.85	0.14	0.00	0.92	0.46	0.49
9	0.54	0.35	0.95	0.57	0.68	0.00	0.55	0.56
10	0.27	0.00	0.75	0.29	0.43	0.78	0.38	0.45
11	0.00	0.74	0.86	0.65	0.00	0.41	0.49	0.48
12	0.13	0.56	0.00	0.00	0.00	0.99	0.24	0.30
13	0.40	0.72	0.77	0.28	0.00	0.73	0.51	0.52
14	0.92	0.98	0.90	0.71	0.64	0.60	0.83	0.85

6 结论(Conclusion)

本研究针对教师课堂教学视频，建立全自动评价教师教学行为的框架。这个框架基于人体姿势识别模型HRNet获得了教师的鼻子、左手和右手的关键点信息，在这些关键点信息的基础上进一步建立描述教师教学行为状态的评价指标，将这些评价指标依据模糊评价综合算法，获得了教学视频的综合评分。在14个教学演示视频中，证实了本框架是可行的，为实现教师全教学过程的智能化分析和自动化评价提供了有益的参考。而且本算法框架具有较好的鲁棒性，例如关键点概率阈值的取值范围比较广。第3节中取阈值是0.7，而

实际上，阈值取值为[0.6,0.8]并不影响最后的综合评价结果，只是个别评价指标数值略有变化。

然而教师教学的视频行为分析是一个新颖的研究领域，还有很多值得深入研究的问题，主要包括：

(1)目前本研究的教学视频是教师在黑板前的教学行为，而教师不在黑板前的教学行为需要进一步深入研究；

(2)目前本研究是完全依据视频进行分析，没有结合教师的授课语音综合分析，教师授课语音和教学行为的信息融合分析有待突破，评价指标需要进一步增加；

(3)教师的教学视频样本量还不够大，今后需要建立一个覆盖各学段、各学科、各年级，以及不同教师的教龄、职称、学历等方面的教师教学视频库。

参考文献(References)

- [1] DOHMS M C, COLLARES C F, TIBERIO I C. Video-based feedback using real consultations for a formative assessment in communication skills[J]. BMC Med Educ, 2020, 20(1):1-9.
- [2] HUANG W, LI N, QIU Z J, et al. An automatic recognition method for students' classroom behaviors based on image processing[J]. Trait Signal, 2020, 37(3):503-509.
- [3] 闫晓炜,张朝晖,赵小燕,等.用于课堂教学评估的教师肢体动作检测[J].中国教育信息化,2019,16(8):88-91.
- [4] 周鹏霄,邓伟,郭培育,等.课堂教学视频中的S-T行为智能识别研究[J].现代教育技术,2018,28(06):54-59.
- [5] BLACKWELL J. Differentiating between teaching experience and expertise in the music studio: A pilot study[J]. Res Stud Music Educ, 2020, 14(4):1-13.
- [6] GREVE S, WEBER K E, BRANDES B, et al. Development of pre-service teachers' teaching performance in physical education during a long-term internship analysis of classroom videos using the Classroom Assessment Scoring System K-3[J]. Ger J Exerc Sport Res, 2020, 11(3):343-353.
- [7] XU X X, ZOU Q, LIN X, et al. Integral knowledge distillation for multi-person pose estimation[J]. IEEE Signal Process Lett, 2020, 27(2):436-440.
- [8] SUN K, XIAO B, LIU D, et al. Deep high-resolution representation learning for human pose estimation[C]. 2019 IEEE Conference on Computer Vision and Pattern Recognition. Long Beach, 2019:5686-5696.
- [9] 李鹏.电视节目主持人的手势语浅析[J].中国广播电视学刊,2019(08):100-102.
- [10] 何梅.基于模糊综合评价法的本科教学评价[J].中国多媒体与网络教学学报(上旬刊),2020(07):34-36.

作者简介:

郑誉煌(1979-),男,博士,副教授.研究领域:机械电子工程.