

慕课授课中的学生听课行为自动分析系统

戴亚平^{1,2} 杨方方^{1,2} 赵翰奕³ 贾之阳^{1,2} 广田熏^{1,2}

摘要 为了解决在线课程 (Massive open online course, MOOC) 授课过程中, 缺乏对于学生学习情况的跟踪与教学效果评估问题, 本文依据视频信息对学生行为进行建模, 提出了一种评判学生听课专心程度的行为自动分析算法. 该算法能够有效跟踪学生的学习状态, 提取学生的行为特征参数, 并对这些参数进行 D-S 融合判决, 以获得学生的听课专注度. 经过多次实验的结果表明, 本文采用的方法能够有效评判学生在授课期间的专心程度, 在数据融合上, 与贝叶斯推理方法相比, 采用 D-S 融合方法能有效提高实验结果的准确性和可靠性.

关键词 学生注意力建模, 特征提取, 决策融合算法, 慕课

引用格式 戴亚平, 杨方方, 赵翰奕, 贾之阳, 广田熏. 慕课授课中的学生听课行为自动分析系统. 自动化学报, 2020, 46(4): 681–694

DOI 10.16383/j.aas.c170416

Auto Analysis System of Students Behavior in MOOC Teaching

DAI Ya-Ping^{1,2} YANG Fang-Fang^{1,2} ZHAO Han-Yi³ JIA Zhi-Yang^{1,2} HIROTA Kaoru^{1,2}

Abstract Aiming at solving the problems of students learning behavior tracking and instructors teaching evaluation in massive open online course (MOOC), a modeling approach of student attention is proposed first, then an automatic behavior analysis and decision making fusion algorithm (ABA) is proposed to evaluate the concentration of the students during lectures. The proposed method can effectively track the student's learning state and acquire the characteristic parameters of the student, and then give the concentration evaluation of the student after data fusion and decision making. Multiple experiments are carried out using the approach proposed in this paper, the results show that the proposed method can effectively reduce the uncertainty in student behavior decision making.

Key words Student attention modeling, feature extraction, decision fusion, massive open online course (MOOC)

Citation Dai Ya-Ping, Yang Fang-Fang, Zhao Han-Yi, Jia Zhi-Yang, Hirota Kaoru. Auto analysis system of students behavior in MOOC teaching. *Acta Automatica Sinica*, 2020, 46(4): 681–694

在信息爆炸的“互联网+”新时代, 随着网络时代的发展, 人类的注意力开始变短, “碎片化”学习成为人类的主要学习模式, 而在线网络课程由于其“微课”的教学模式, 时间的相对自由性, 成本相对较低等优点迅速受到广大学者的喜爱^[1-4]. 伴随着普通学生、成年人、工作者越来越多地进行在线网络学习, 许多机构也采用网络课程模式对其员工进行培训, 但是不同于传统的培训, 在线网络学习难以被管控, 机构无法得到员工的培训效果, 因此, 即使意识到这种学习的重要性, 许多企业也不知道如何“培育”它. 这是因为在线网络课程目前仍处于完

善阶段, 其本身存在一些缺陷, 由于时间的自由和目前的无人监督模式, 学生的学习效率难以得到保障, 学生很可能因为自律性不够或课程的枯燥而出现不专心的情况^[5]. 针对这种情况, 有必要对学习者的学习状态进行分析和评估, 以监督学生的学习状态, 给相关机构提供测评依据, 也反馈给在线授课者, 帮助教育者提高授课质量^[6].

目前在已有的研究成果中很难找到针对 MOOC (Massive open online course) 授课过程中学生的行为检测, 虽然在其他应用背景下的人体行为检测已经有了很多, 但是针对 MOOC 授课过程中的学生这一特定目标, 对其行为进行检测对提升 MOOC 的教育教学还是有不容小觑的意义. 为了获取学生的专注度, 首先要做的就是获得学生的行为信息, 考虑到学生登录慕课时, 所使用的电脑、平板和手机等设备一般都带有摄像头, 如果能够利用摄像头来实时记录学生的学习状态, 就可以根据获取到的视频对学生的行为进行分析, 以得到学生的学习专注度. 目前国内外的学者在行为识别领域上已

收稿日期 2017-07-26 录用日期 2018-07-15
Manuscript received July 26, 2017; Accepted July 15, 2018
本文责任编辑 朱军

Recommended by Associate Editor ZHU Jun

1. 北京理工大学自动化学院 北京 100081 2. 复杂系统智能控制与决策国家重点实验室 北京 100081 3. 华北计算机系统工程研究所 (信息产业部电子第六研究所) 北京 100081

1. School of Automation, Beijing Institute of Technology, Beijing 100081 2. National Laboratory of Intelligent Control and Decision of Complex Systems, Beijing 100081 3. The 6th Research Institute of China Electronics Corporation, Beijing 100081

经取得了一系列成果, Jaouedi 等为了更好地使机器人服务于残疾人, 他们对残疾人与智能机器人交流中的手势、步行速率等信息进行了分析, 之后使用卡尔曼滤波方法对移动的人体进行跟踪, 最后采用“最近邻居法”对所得到的行为信息进行处理, 以给出人体的行为决策^[7]; 而 Xiao 等对人机交互中的头部运动进行了行为分析研究, 他们提出了以数据为基础来识别头部运动模式的方法, 先将人体头部运动按照时间进行分段, 再用线性预测特征参数法来建立运动模型, 之后用高斯混合模型对运动类型进行融合推理, 得出头部运动所代表的特定行为, 例如接受、赞赏、责备、反对等^[8]; Tsai 等为了检测公共场所人类吸烟饮酒等非法行为, 他们将相机安装在工作场所的某一固定位置, 然后提取三个主要特征: 脸与手的接触时间特征、烟雾特征和手持物体特征, 最后根据“决策树”来判断是否有人在公共场所从事了吸烟饮酒等非法行为^[9].

虽然目前对人体行为的检测已经取得了较多成果, 并且可以实现人体的跟踪问题^[7-11], 但是利用学生听课过程中的视频来“判定学生在线网络学习中的专注度”这一方面的研究还比较空缺^[12]. 针对慕课这一新的应用环境, 我们分析摄像头获取的学生行为信息, 建立人体的“四维特征信息模型”, 根据各个特征的变化范围对每个焦元进行基本概率赋值, 最后采用 D-S 方法融合时空域中的概率值, 判断出学生的人体信息, 最后给出学生对某一门慕课课程学习效果的初步评价.

该分析系统可以用图 1 来表示.

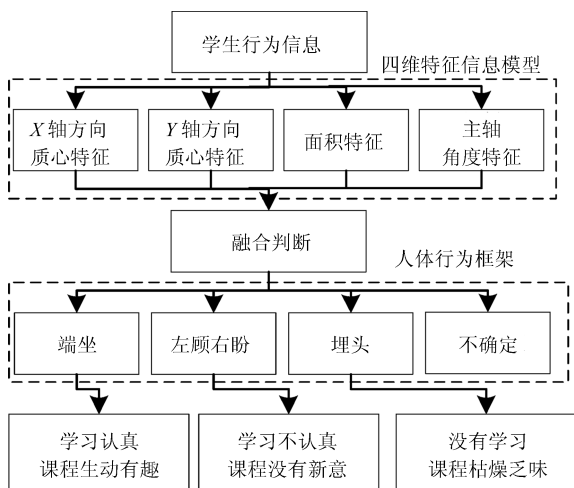


图 1 慕课授课过程中学生专注度自动检测与判定系统
Fig. 1 Automatic detection system on student focus during MOOC teaching

论文的主要工作组成共分为以下内容: 第 1 节, 提取人体轮廓, 并建立学生“四维特征信息模型”; 第 2 节, 搭建人体行为框架, 并对其进行基本概率赋

值, 之后采用 D-S 方法对其进行融合, 最后给出学生的行为状态; 第 3 节, 采用多种状态的视频信息对以上算法进行了实验验证, 并将 D-S 理论与贝叶斯理论进行对比, 来证明 D-S 理论融合的有效性; 第 4 节, 总结和展望.

1 慕课授课过程中学生行为模型的建立

1.1 慕课授课过程中学生人体检测

慕课授课过程中的运动人体检测任务是将人体从视频序列中分割出来, 因此, 能否准确地提取出人体轮廓就显得尤为重要^[13-15].

本文采用的是背景减除法提取目标人体^[16], 由于实验地点相对单一, 我们使用无学生的背景图片作为背景模型, 将视频中当前帧图像与背景图像进行差分运算, 如果像素点灰度值差别很大, 则为目标人体, 如果差值在一定阈值范围内, 则为背景像素点; 得到的差值图像采用最大类间方差法找到合适的二值化阈值; 再对二值化后的图像提取出最大连通域, 消除背景的干扰, 最后我们还使用基于逆合成孔径雷达 (Inverse synthetic aperture radar, ISAR) 图像目标形状提取的数学形态学方法对检测到的前景目标进行后处理^[17], 以平滑连接目标人体的边缘, 使提取出的人体轮廓更加完整.



(a) 原始图像
(a) Original image



(b) 提取出的人体目标
(b) Extracted human body

图 2 电脑摄像头提取的目标人体图像

Fig. 2 Human body image obtained by computer camera

1.2 建立“四维特征信息模型”

提取人体轮廓后, 还需进一步提取出图像的主要特征信息, 才能进行特征信息融合与听课专注度

的决策推理. 根据学生在学习时可能出现的行为状态, 例如认真学习时的端坐状态, 有点疲惫时的斜坐状态, 非常困倦时可能会出现趴在桌子上睡觉的状态等, 为了能全面且准确地描述这些人体的行为信息, 我们从人体轮廓中提取了 4 个行为特征信息, 将其定义为人体“四维特征信息模型”, 其具体内容:

$$\left\{ \begin{array}{l} \text{人体 } x \text{ 轴质心特征: } X = \frac{M_{10}}{M_{00}} \\ \text{人体 } y \text{ 轴质心特征: } Y = \frac{M_{01}}{M_{00}} \\ \text{面积特征标记矩阵:} \\ \quad \text{STATS} = \text{regionprops}(L, \text{properties}) \\ \text{人体主轴特征: } \theta = \text{atan2}(2b, a - c + \text{sqrt}) \end{array} \right. \quad (1)$$

其中, *regionprops* 和 *atan2* 分别为 Matlab 中度量图像区域属性和四象限反正切函数. 我们通过判断学生人体 *x* 轴质心特征, 人体 *y* 轴质心特征和人体主轴特征是否偏离原点位置来判定其是否处于端坐状态, 如果处于端坐状态, 则记录端坐时的面积特征, 并将端坐时的四维特征作为基准, 当学生四维特征信息出现变化时, 再去判断其处于哪种状态. 从目标提取中, 我们得到了最大连通域, 然后使用基于空间分布特征的多连通域融合算法来计算四维特征信息中的人体的质心位置, 人体的面积和人体主轴信息^[18].

1) 人体质心特征 (*X*, *Y*)

对于质心, 可以认为得到的二值图像或者灰度图像是一个二维的密度分布函数. 对于一个二维连续函数 $f(x, y)$, 其零阶矩 M_{00} 和一阶矩 M_{10} 、 M_{01} 分别为

$$\begin{aligned} M_{00} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy \\ M_{10} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f(x, y) dx dy \\ M_{01} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f(x, y) dx dy \end{aligned} \quad (2)$$

则得到人体的质心为:

$$X = \frac{M_{10}}{M_{00}}, \quad Y = \frac{M_{01}}{M_{00}} \quad (3)$$

其中, (*X*, *Y*) 为目标人体的质心坐标.

2) 学生人体主轴方向变化特征 θ

对中心矩进行归一化操作:

$$\mu_{pq} = N_{pq} = \frac{M_{pq}}{M_{00}^r}, \quad r = \frac{p+q}{2} + 1 \quad (4)$$

然后利用:

$$f(x) = \begin{cases} a = \frac{\mu_{20}}{\mu_{00}} \\ b = \frac{\mu_{11}}{\mu_{00}} \\ c = \frac{\mu_{02}}{\mu_{00}} \end{cases} \quad (5)$$

得到

$$\begin{cases} \text{sqrt} = \sqrt{4b^2 - (a - c)^2} \\ \theta = \text{atan2}(2b, a - c + \text{sqrt}) \end{cases} \quad (6)$$

其中, θ 是学生人体主轴方向与垂直方向的夹角, 以作为人体主轴变化特征.

3) 面积特征 *S*

将得到的连通域进行标记, 而标记的方法有很多种, 有的算法可以一次遍历图像完成标记, 有的则需要 2 次或更多次遍历图像, 这样也就造成了不同的算法效率的差异.

我们使用 *bwlabel* 和 *regionprops* 算子对连通域进行操作. *bwlabel* 算子是二值图像中的标签连接组件, 式 (7) 返回一个标签矩阵, 其中变量 *n* 表示的是指定的连接方向数量, 而 *num* 表示的是在图像 *BW* 中发现的连接对象的数量.

$$[L, \text{num}] = \text{bwlabel}(BW, n) \quad (7)$$

而 *regionprops* 算子是一个度量图像区域属性的函数, 式 (8) 返回的是标签矩阵 *L* 中每个标记区域的特征值.

$$\text{STATS} = \text{regionprops}(L, \text{properties}) \quad (8)$$

在得到的标记矩阵 *STATS* 中, 不同的数字代表着不同的区域, 面积标记矩阵如式 (9) 所示, 然后累加最大域中的值就可以得到人体面积特征 *S*.

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (9)$$

2 基于改进 D-S 方法的学生注意力模型分析与决策

随着网络时代的到来, 人们生活中信息的形式、数量、复杂度都大大增加, 因此对信息处理的及时性、准确性的要求也越来越高. 数据融合是指通过对多个同质或异质传感器获得的数据进行有机结合, 最后给出检测结果. 其实信息融合存在于我们生活中的方方面面, 例如人类将各种感觉器官(眼、耳、鼻)所探测到的(图像、声音、气味)信息与先验知识进行综合, 得出对周围的环境综合判断, 以决定接下来的行动^[19]. 本文在信息融合时选用改进的 D-S 算法来处理得到的信息, 先获取人体四维特征信息, 之后采用“基本概率赋值法”对每个焦元进行概率赋值, 最后对特征信息进行融合, 从而计算出学生人体行为框架中每个状态的概率分配^[20].

2.1 人体行为框架的建立及基本概率赋值

根据学生在线网络学习时可能出现的行为状态, 结合专家经验和理论常识, 我们将人体识别框架设定为 $\Omega = \{ \text{左顾右盼, 埋头, 端坐, 不确定} \}$ 4 种行为状态, 分别用 h_1, h_2, h_3 和 U 来表示, 3 种人体行为如图 3 所示.



图 3 电脑获取的人体图像的三种形态

Fig. 3 Three forms of human body image acquired by computer

接下来为每一个证据体进行概率赋值, 基本概率赋值是相关领域专家依据自身经验和已有知识对整个识别框架中不同证据进行不同归纳判断而得到的, 其中含有很强个人主观性, 所以不同专家对于同一个证据不可能给出完全一样的基本概率分配, 有时差别会很大.

在本文中我们将人体的 X 轴质心变化, Y 轴质心变化, 面积变化和角度变化作为 4 个独立的证据体: $M = \{X, Y, S, \theta\}$, 用 X, Y, S, θ 分别表示当前时刻目标人体的 X 轴坐标值, Y 轴坐标值, 面积和主轴角度, 当人体行为出现变化时, 为了定性地给 4 个证据体赋值, 我们在每个证据体的曲线中分别选择两个阈值, 分别计算 4 个证据体 (X, Y, S, θ) 每条曲线在阈值外和阈值内所占的百分比, 像图 4 所示的那样, 再使用式 (10)~式 (14), 就可以得到人体辨识框架中每一部分的基本概率赋值.

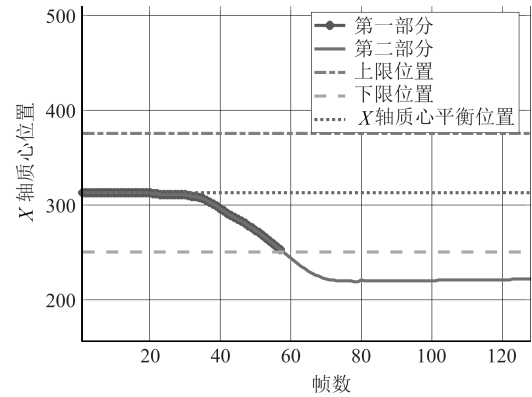


图 4 X 轴质心坐标变化曲线的两部分

Fig. 4 Two parts of X axis centroid position curve

当人体处于端坐状态时, 4 个证据体的数值基本保持不变; 当人体左顾右盼时, 人体的质心坐标和方向会不断发生变化; 当人埋头时, 人体的 Y 轴质心坐标会发生变化, 同时人体的面积也会显著减小. 结合学生在授课时的行为状态, 本文提出一种新的基本概率赋值方法—四维基本概率赋值法, 该方法不仅充分利用了 4 种证据体所包含的信息, 也在很大程度上提高了 D-S 决策的准确率, 同时又避免了分析单一证据所得结果的不完整性和片面性.

四维基本概率赋值公式如下所示, 其中用 $X_{threshold1}$ 和 $X_{threshold2}$ 分别表示 X 轴质心变化的上下阈值, Y, S, θ 的上下阈值依次类推, $S_{threshold3}$ 为一个很小的值, 代表学生不在画面中时人体的面积, 为防止光照, 背景等因素的干扰, 这里不设定为 0, 而设定为一个很小的值.

$$\begin{cases} \text{若 } X > X_{threshold1} \text{ 或 } X < X_{threshold2}, \\ \quad \text{左顾右盼的次数} + 1 \\ \text{否则, 埋头的次数} + 0.5, \quad \text{端坐的次数} + 0.5 \end{cases} \quad (10)$$

$$\begin{cases} \text{若 } Y > Y_{threshold1} \text{ 或 } Y < Y_{threshold2}, \\ \quad \text{埋头次数} + 1 \\ \text{否则, 左顾右盼次数} + 0.5, \quad \text{端坐的次数} + 0.5 \end{cases} \quad (11)$$

$$\begin{cases} \text{若 } S > S_{threshold1} \text{ 或 } S < S_{threshold2}, \\ \quad \text{埋头次数} + 1 \\ \text{否则, 左顾右盼次数} + 0.5 \quad \text{端坐次数} + 0.5 \end{cases} \quad (12)$$

$$\begin{cases} \text{若 } \theta > \theta_{threshold1} \text{ 或 } \theta < \theta_{threshold2}, \\ \quad \text{左顾右盼次数} + 0.5, \quad \text{埋头次数} + 0.5 \\ \text{否则, 端坐次数} + 1 \end{cases} \quad (13)$$

$$\left\{ \begin{array}{l} \text{若 } (X > X_{threshold1} \text{ 或 } X < X_{threshold2}) \\ \text{且 } (Y > Y_{threshold1} \text{ 或 } Y < Y_{threshold2}) \\ \text{且 } (S > S_{threshold1} \text{ 或 } S < S_{threshold2}) \\ \text{且 } (\theta > \theta_{threshold1} \text{ 或 } \theta < \theta_{threshold2}) \\ \text{或 } (S < S_{threshold3}), \\ \text{不确定次数} + 1 \end{array} \right. \quad (14)$$

2.2 D-S 人体信息融合模型

D-S 证据理论是关于证据和可能性推理的理论, 它主要处理证据加权和证据支持度问题, 并且利用可能性推理来实现证据的组合. 随着信息融合的发展, 我们融合的对象不再局限于传感器获取的数据和信息, 而是可推广到各种方法, 专家经验甚至思想等.

我们将从实验中得到的学生的学习视频一共分为 n 个部分, 针对实验过程中收集到的 4 个证据体 $\{X, Y, S, \theta\}$, 首先计算当前帧中每个证据体的概率值, 即判断 4 个证据体的变化是否有超出设定的阈值, 得到每一帧中每个证据体对每种行为框架 (端坐、左顾右盼、埋头和不确定) 的概率分配值, 在实验中 $m(h_{1jk}), m(h_{2jk}), m(h_{3jk}), m(h_{4jk})$ 分别表示 X 轴质心坐标变化, Y 轴质心坐标变化, 面积变化以及角度变化 4 个证据体的值, j 表示第 j 个人体框架 ($1 \leq j \leq m$), k 表示第 k 个周期 ($1 \leq k \leq n$).

然后将每个证据体对每种框架的概率值进行空间域上的融合, 即每个周期中传感器之间的融合. 在这里是指融合 4 个证据体对每种行为框架概率值, 就可以得到在每个周期中 4 种人体框架的概率值, 如式 (15) 所示:

$$m(h_{jk}) = \frac{1}{K} \sum_{h_{1jk} \cap \dots \cap h_{4jk} = h_{jk}} m(h_{1jk})m(h_{2jk})m(h_{3jk})m(h_{4jk}) \quad (15)$$

其中, K 是归一化因子.

$$K = \sum_{h_{1jk} \cap \dots \cap h_{4jk} \neq \emptyset} m(h_{1jk})m(h_{2jk})m(h_{3jk})m(h_{4jk}) = 1 - \sum_{h_{1jk} \cap \dots \cap h_{4jk} = \emptyset} m(h_{1jk})m(h_{2jk})m(h_{3jk})m(h_{4jk}) \quad (16)$$

最后将各个时间段内 4 种框架的概率值进行时间域上的融合, 即进行周期之间的融合, 得到最后的时空融合结果. 也就得到了最终的 4 种坐姿的概率分配 $m(k)$, 经过比较大小就可以判断出在这个时间

段中人体的坐姿情况, 计算过程如式 (17) 所示:

$$m(h_j) = m(h_{j1}) \oplus m(h_{j2}) \oplus \dots \oplus m(h_{jk}) = \frac{1}{K} \sum_{h_{j1} \cap h_{j2} \cap \dots \cap h_{jn} = h_j} m(h_{j1})m(h_{j2}) \dots m(h_{jn}) \quad (17)$$

同理, K 是归一化因子.

在本文中取 $m = 4, n = 3$, 最后得到的融合结果是学生在听课这一段中的行为评价结果, 可以用该结果作为学生学习状态评价的一个初步依据.

智能教室中学生行为分析与推理决策模型的实验流程图如图 5 所示.

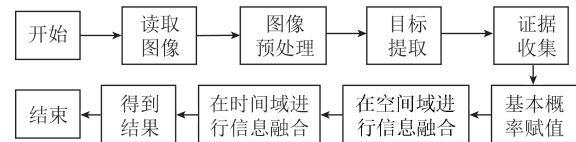


图 5 学生行为分析与推理决策流程图

Fig. 5 Flow chart of student behavior analysis and decision-making

3 实验结果和分析

在实验中, 我们使用笔记本电脑的摄像头对学生的行为进行录像跟踪, 之后对视频进行图像处理得到 4 个证据体的关键信息, 然后使用 D-S 证据理论对信息融合后决策. 实验中的电脑使用 Inter Core i7 4720 处理器和 8 GB 内存, 摄像头参数为: 尺寸大小 640×480 , 采样频率为 30 fps, 使用的软件环境为 MATLAB 2016a.

首先我们对三种典型的行为进行了实验, 并且将实验中得到的视频按照上面所说的方法分为三部分, 先在每一部分进行空间域的融合, 得到的结果再进行时间域的融合以获得最终的结果.

3.1 学生端坐状态实验

在这个实验中, 我们模拟了学生专心学习的场景, 学生大部分时间处于端坐状态, 我们得到视频中的关键帧如图 6 所示. 同时得到实验的 4 条曲线如图 7.



图 6 端坐状态时关键帧的图像

Fig. 6 Image of key frame in sit up condition

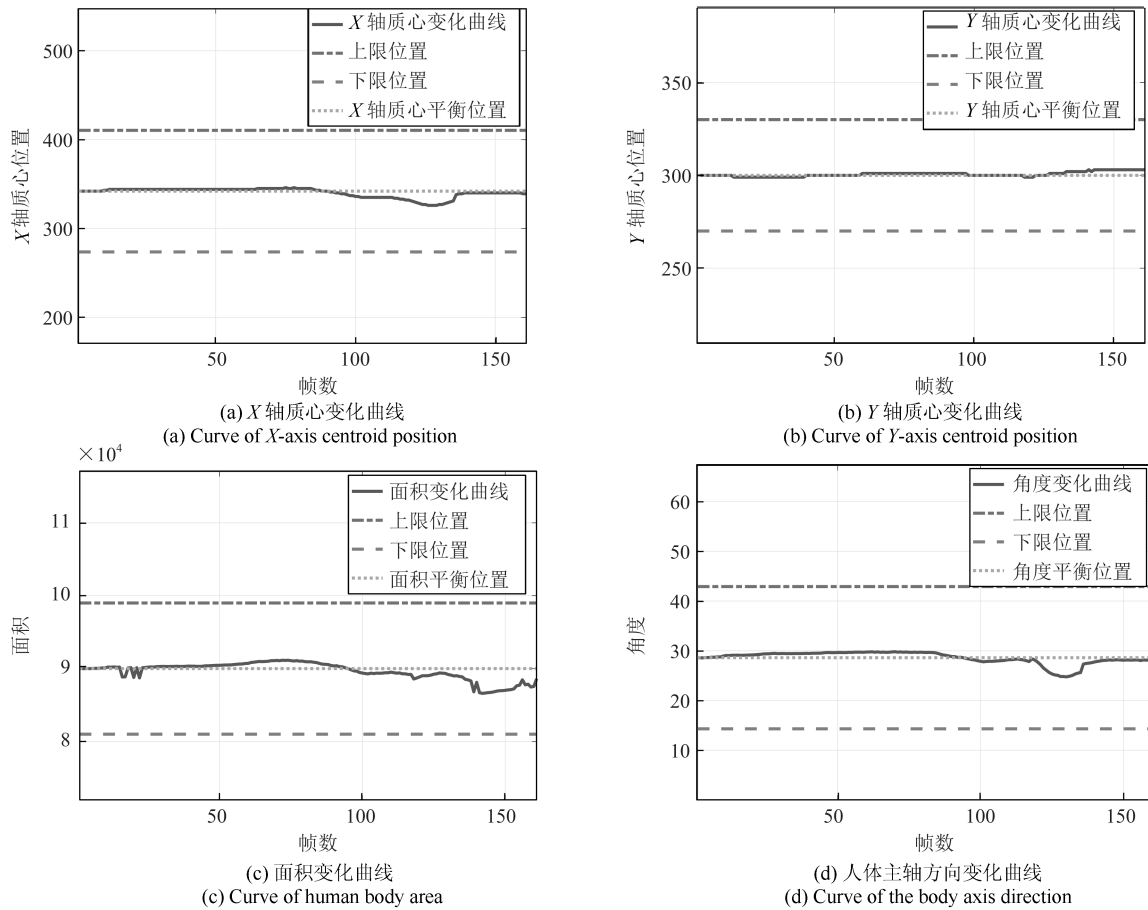


图 7 端坐状态特征变化曲线
Fig. 7 Characteristic curves of "sit up" state

表 1 端坐时的基本概率赋值

Table 1 The probability distributions when sit up

	h_1	h_2	h_3	U
$m(h_{1j1})$	0	0.4444	0.4444	0.1112
$m(h_{1j2})$	0	0.4528	0.4528	0.0944
$m(h_{1j3})$	0	0.4528	0.4528	0.0944
$m(h_{2j1})$	0.4444	0	0.4444	0.1112
$m(h_{2j2})$	0.4528	0	0.4528	0.0944
$m(h_{2j3})$	0.4528	0	0.4528	0.0944
$m(h_{3j1})$	0.4444	0	0.4444	0.1112
$m(h_{3j2})$	0.4528	0	0.4528	0.0944
$m(h_{3j3})$	0.4528	0	0.4528	0.0944
$m(h_{4j1})$	0	0	0.8888	0.1112
$m(h_{4j2})$	0	0	0.9056	0.0944
$m(h_{4j3})$	0	0	0.9056	0.0944

计算每一部分的基本概率赋值我们可以得到表 1, 其中 $m(h_{ijk})$ 代表第 i 个证据体 (X 轴质心特征, Y 轴质心特征, 面积特征, 角度特征) 在第 k 个周期 ($k = 1, 2, 3$) 对第 j 个人体识别框架 (左顾右盼、埋头、端坐和不确定) 的基本概率赋值. 将表 1 中学生端坐时各个阶段每个证据体对每种人体行为框架的

基本概率值进行空间域的融合可以得到表 2, 即得到了三个周期中每种人体行为框架的概率值.

表 2 空间域融合后的结果

Table 2 The results of fusion in the spatial domain

	h_1	h_2	h_3	U
$m(h_{j1})$	0.0191	0.0036	0.9773	0
$m(h_{j2})$	0.0142	0.0023	0.9834	0
$m(h_{j3})$	0.0142	0.0023	0.9834	0

之后将表 2 中三个周期每个证据体的概率值进行时间域上的融合, 即进行周期之间的融合, 得到 $m(h_1) = 4.1001 \times 10^{-6}$, $m(h_2) = 2.1133 \times 10^{-8}$, $m(h_3) = 0.9999$, $m(U) = 4.7766 \times 10^{-24}$, 已知 $m(h_3) > m(h_1) > m(h_2) > m(U)$, 所以最终的决策结果是 h_3 . 即学生在这段时间基本处于端坐状态, 这和我们的认知是符合的. 也说明课程相对比较轻松有趣, 学生学习比较认真, 授课者需要继续保持这种授课模式, 相关机构可以认为员工在线培训比较成功.

为了验证“四维特征信息模型”提取的有效性, 在本次实验中我们还做了三组对比实验. 实验 1 采

用 X 轴质心变化和 Y 轴质心变化二维特征来判断学生的专注度, 它作为本实验的 *baseline*; 实验 2 采用 X 轴质心变化, Y 轴质心变化和面积变化来判断学生在这段时间的学习行为; 实验 3 采用本文提出的“四维特征信息模型”对学生的行为进行判断. 实验结果如表 3 所示.

表 3 多维特征信息融合结果对比

Table 3 Comparison of multi-dimensional feature information fusion results

实验方法	实验结果			
	$m(h_1)$	$m(h_2)$	$m(h_3)$	$m(U)$
$X + Y$ (<i>baseline</i>)	0.0042	0.0042	0.9916	0
$X + Y + S$	0.0042	0	0.9958	0
$X + Y + S + \theta$	0	0	0.9999	0

从实验结果可以看出, 采用“四维特征信息模型”可以有效提高结果的准确率, 减少其他状态的干扰, 使结果更具有信服力.

3.2 学生左顾右盼状态实验

在这个实验中, 学生一直在左顾右盼, 我们得到

视频中的关键帧如图 8 所示.



图 8 左顾右盼时关键帧的图像

Fig. 8 Image of key frame in look around condition

我们同时得到学生人体特征变化的四条曲线如图 9.

和端坐时一样, 计算每一部分的基本概率赋值我们可以得到该实验的基本概率赋值如表 4.

同理, 将表 4 中学生左顾右盼时的基本概率值进行空间域的融合可以得到表 5.

之后进行时间域上的融合, 得到 $m(h_1) = 0.9013, m(h_2) = 7.7065 \times 10^{-5}, m(h_3) = 0.0987, m(U) = 4.4837 \times 10^{-21}$, 已知 $m(h_1) > m(h_3) > m(h_2) > m(U)$, 所以最终的决策结果是 h_1 , 即学生在这段时间经常左右摇晃, 可以认为学生学习不认真, 可能是这节慕课课程没有太大的

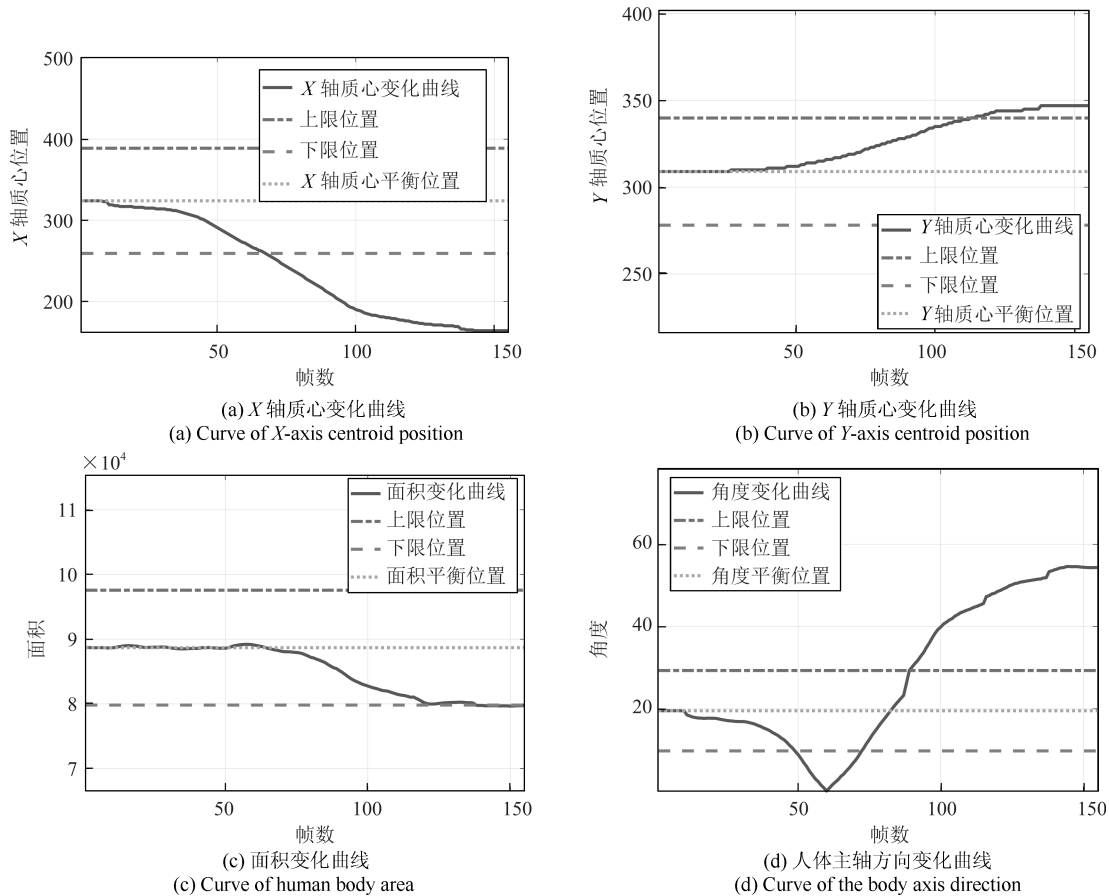


图 9 左顾右盼状态特征变化曲线

Fig. 9 Characteristic curves of “look around” state

表 4 左顾右盼状态的基本概率赋值

Table 4 The probability distributions when look around

	h_1	h_2	h_3	U
$m(h_{1j1})$	0	0.4442	0.4442	0.1116
$m(h_{1j2})$	0.6271	0.1394	0.1394	0.0942
$m(h_{1j3})$	0.9058	0	0	0.0942
$m(h_{2j1})$	0.4442	0	0.4442	0.1116
$m(h_{2j2})$	0.4529	0	0.4529	0.0942
$m(h_{2j3})$	0.0784	0.7490	0.0784	0.0942
$m(h_{3j1})$	0.4442	0	0.4442	0.1116
$m(h_{3j2})$	0.4529	0	0.4529	0.0942
$m(h_{3j3})$	0.4529	0	0.4529	0.0942
$m(h_{4j1})$	0.0261	0.0261	0.8361	0.1116
$m(h_{4j2})$	0.3135	0.3135	0.2787	0.0942
$m(h_{4j3})$	0.4529	0.4529	0	0.0942

表 5 空间域融合后的结果

Table 5 The results of fusion in the spatial domain

	h_1	h_2	h_3	U
$m(h_{j1})$	0.0249	0.0047	0.9704	0
$m(h_{j2})$	0.7743	0.0047	0.2210	0
$m(h_{j3})$	0.9228	0.0681	0.0091	0

吸引力, 有点枯燥乏味, 授课者应注意改善授课质

量, 也可能是因为学生学习态度不认真, 培训机构可以提醒相应员工经过后期检查决定该员工是否需要重新学习, 以取得良好的学习效果.

3.3 学生埋头状态实验

在这个实验中, 学生处于埋头状态, 我们得到视频中的关键帧如图 10 所示. 我们得到实验的四条曲线如图 11 所示.



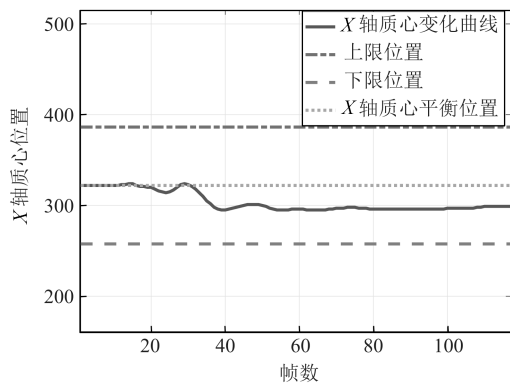
图 10 埋头时关键帧的图像

Fig. 10 Image of key frame in head drop

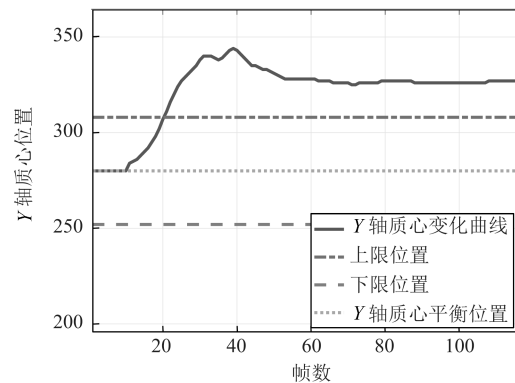
计算每一部分的基本概率赋值我们可以得到该实验的基本概率赋值如表 6.

将表 6 中学生埋头时的基本概率值进行空间域的融合可以得到表 7.

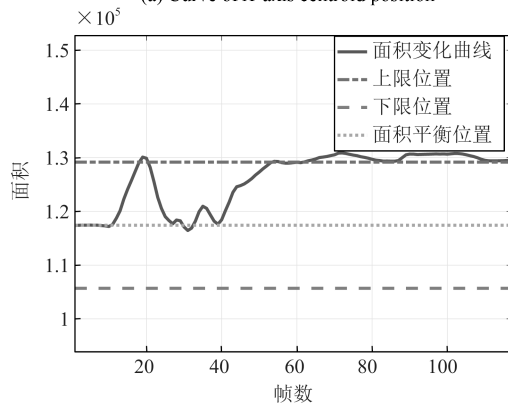
之后进行时间域上的融合, 得到 $m(h_1) = 3.8405 \times 10^{-8}$, $m(h_2) = 0.9360$, $m(h_3) = 0.0640$,



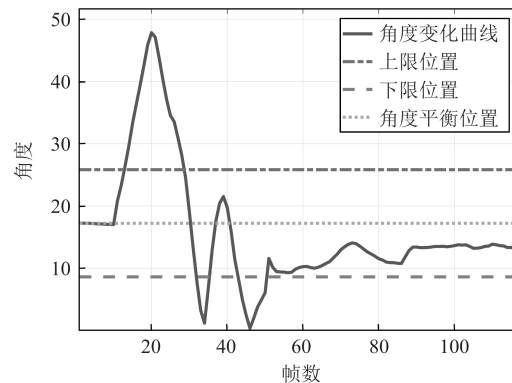
(a) X轴质心变化曲线
(a) Curve of X-axis centroid position



(b) Y轴质心变化曲线化曲线
(b) Curve of Y-axis centroid position



(c) 面积变化曲线
(c) Curve of human body area



(d) 人体主轴方向变化曲线
(d) Curve of the body axis direction

图 11 埋头状态特征变化曲线

Fig. 11 Characteristic curves of "head drop" state

$m(U) = 7.0128 \times 10^{-21}$, 已知 $m(h_2) > m(h_3) > m(h_1) > m(U)$, 所以最终的决策结果是 h_2 . 即学生在这段时间大部分处于埋头状态, 可以推测学生在这段时间并没有听课, 而是爬在桌子上, 可能学生这段时间比较疲惫, 也可能这节慕课课程太枯燥乏味, 慕课平台应着重审核该课程是否符合慕课上线标准, 如果该学生属于个例, 培训机构应要求该学生重新学习这门课程, 以达到培训的目的.

表6 埋头时的基本概率赋值

Table 6 The probability distributions when head drop

	h_1	h_2	h_3	U
$m(h_{1j1})$	0	0.4500	0.4500	0.1000
$m(h_{1j2})$	0	0.4500	0.4500	0.1000
$m(h_{1j3})$	0	0.4500	0.4500	0.1000
$m(h_{2j1})$	0.2308	0.4385	0.2308	0.1000
$m(h_{2j2})$	0	0.9000	0	0.1000
$m(h_{2j3})$	0	0.9000	0	0.1000
$m(h_{3j1})$	0.4269	0.0462	0.4269	0.1000
$m(h_{3j2})$	0.2308	0.4385	0.2308	0.1000
$m(h_{3j3})$	0	0.9000	0	0.1000
$m(h_{4j1})$	0.2308	0.2308	0.4385	0.1000
$m(h_{4j2})$	0.0923	0.0923	0.7154	0.1000
$m(h_{4j3})$	0	0	0.9000	0.1000

表7 空间域融合后的结果

Table 7 The results of fusion in the spatial domain

	h_1	h_2	h_3	U
$m(h_{j1})$	0.0658	0.2048	0.7294	0
$m(h_{j2})$	0.0020	0.8131	0.1848	0
$m(h_{j3})$	0	0.9221	0.0778	0

3.4 复杂状态实验 1

在这个实验中, 我们模拟了学生处于正常学习时的状态, 在这段时间里, 学生一共左顾右盼了 2 次, 埋了 2 次, 但基本上是处于端坐的学习状态. 实验中的关键帧序列如图 12 所示, 我们同样获得曲线图如图 13.

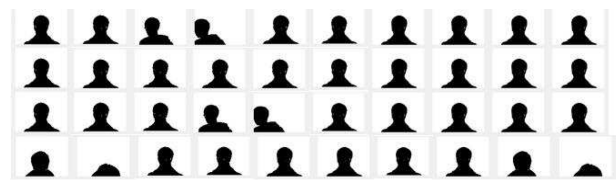


图 12 复杂状态实验 1 的关键帧序列

Fig. 12 Key frame sequence under complex state experiment 1

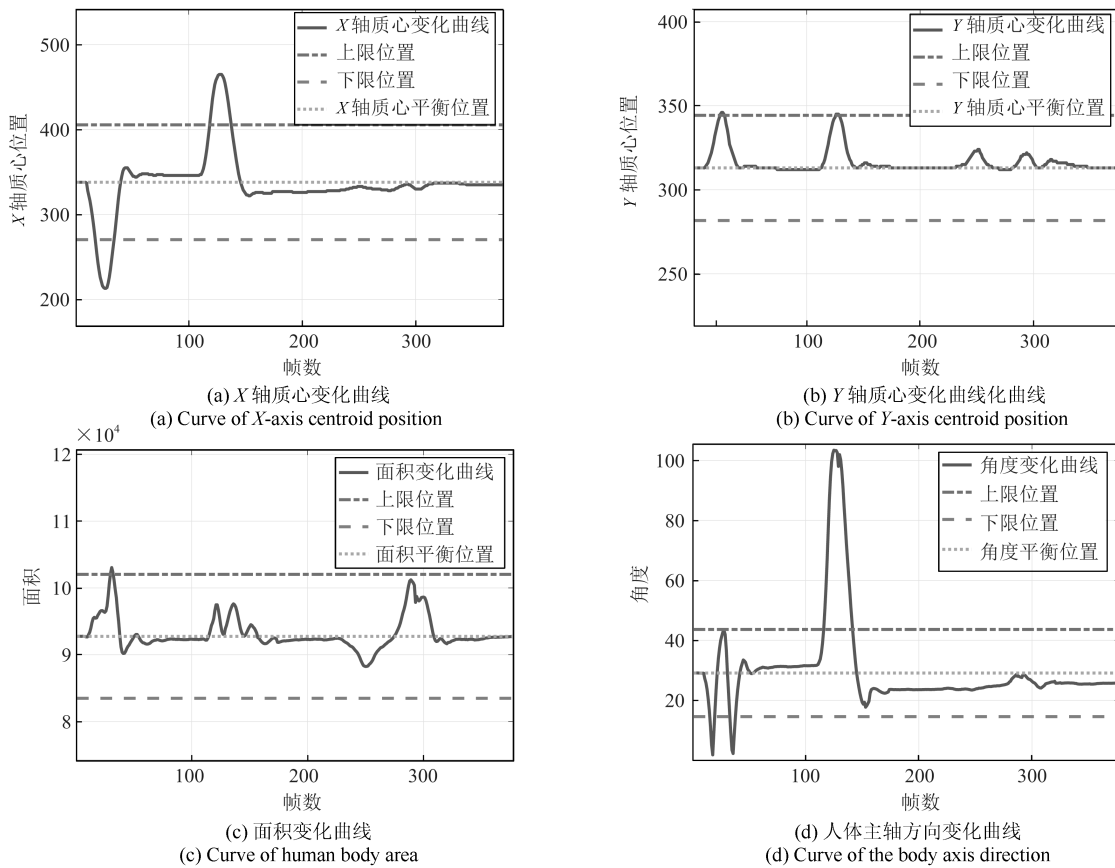


图 13 复杂状态 1 的特征变化曲线

Fig. 13 Characteristic curves of complex state 1

计算每一部分的基本概率赋值我们可以得到该实验的基本概率赋值如表 8.

将表 8 中学生各个状态的基本概率值进行空间域上的融合后,可以得到表 9.

表 8 证据体的基本概率赋值 (复杂状态实验 1)
Table 8 Basic probability distribution of evidence bodies (Complex condition 1)

	h_1	h_2	h_3	U
$m(h_{1j1})$	0.1719	0.3617	0.3617	0.1048
$m(h_{1j2})$	0.0859	0.4082	0.4082	0.0976
$m(h_{1j3})$	0	0	0.4512	0.0976
$m(h_{2j1})$	0.4333	0.0286	0.4333	0.1048
$m(h_{2j2})$	0.4405	0.0215	0.4405	0.0976
$m(h_{2j3})$	0.4512	0	0.4512	0.0976
$m(h_{3j1})$	0.4369	0.0215	0.4369	0.1048
$m(h_{3j2})$	0.4515	0	0.4512	0.0976
$m(h_{3j3})$	0.4512	0.0430	0.4512	0.0976
$m(h_{4j1})$	0.0645	0.0645	0.7663	0.1048
$m(h_{4j2})$	0.0573	0.0573	0.7878	0.0976
$m(h_{4j3})$	0	0	0.9024	0.0976

之后进行时间域上的融合,得到 $m(h_1) = 1.0477 \times 10^{-4}$, $m(h_2) = 1.5324 \times 10^{-7}$, $m(h_3) = 0.9999$, $m(U) = 1.6917 \times 10^{-23}$, 已知 $m(h_3) > m(h_1) > m(h_2) > m(U)$, 所以最终的决策结果是 h_3 . 本次实验模拟了学生处于真实状态下的学习情况,学习时间比较长,里面包括了左顾右盼、埋头、端坐等各种姿态,最终结论学生处于端坐状态,即在这段时间内学生虽然有几次走神情况,但基本上在专心学习,可以得出学生在这段时间学习比较认真的结论.

表 9 空间域融合后的结果 (复杂状态实验 1)
Table 9 Results of fusion in the spatial domain (Complex condition 1)

	h_1	h_2	h_3	U
$m(h_{j1})$	0.1003	0.0087	0.8910	0
$m(h_{j2})$	0.0567	0.0056	0.9377	0
$m(h_{j3})$	0.0151	0.0026	0.9823	0

3.5 复杂状态实验 2

在这个实验中,我们模拟了学习时的另一种情况,在这段时间里,学生长时间的处于埋头的情况,并且有时身体处于倾斜状态.实验的关键帧序列如图 14 所示,我们同样获得曲线图如图 15.

计算每一部分的基本概率赋值可以得到该实验的基本概率赋值如表 10. 将表 10 中学生各个状态的基本概率值进行空间域上的融合后,得到表 11.

再进行时间域上的融合,得到 $m(h_1) = 8.3532 \times 10^{-5}$, $m(h_2) = 0.9863$, $m(h_3) =$

0.0136 , $m(U) = 1.0782 \times 10^{-21}$, 已知 $m(h_2) > m(h_3) > m(h_1) > m(U)$, 所以最终的决策结果是 h_2 . 本次复杂实验 2 模拟了学生处于疲劳状态下的学习情况,里面包括了左顾右盼、埋头、端坐等各种姿态,最终结论是学生处于埋头状态.即学生在这段时间虽然有学习的时候,但总体来看,大部分时间处于埋头状态,所以认定该学生没有认真学习.

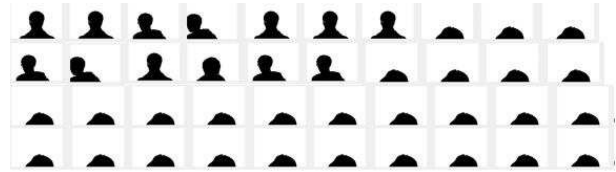


图 14 复杂状态实验 2 的关键帧序列

Fig. 14 Key frame sequence under complex state Experiment 2

表 10 证据体的基本概率赋值 (复杂状态实验 2)
Table 10 Basic probability distribution of evidence bodies (Complex condition 2)

	h_1	h_2	h_3	U
$m(h_{1j1})$	0.3027	0.2951	0.2951	0.1027
$m(h_{1j2})$	0	0.4487	0.4487	0.1027
$m(h_{1j3})$	0	0.4527	0.4527	0.0946
$m(h_{2j1})$	0.4487	0	0.4487	0.1027
$m(h_{2j2})$	0.1415	0.6144	0.1455	0.1027
$m(h_{2j3})$	0	0.9054	0	0.0946
$m(h_{3j1})$	0.4487	0	0.4487	0.1027
$m(h_{3j2})$	0.1900	0.5174	0.1900	0.1027
$m(h_{3j3})$	0	0.9054	0	0.0946
$m(h_{4j1})$	0.3557	0.3557	0.1859	0.1027
$m(h_{4j2})$	0.2425	0.2425	0.4123	0.1027
$m(h_{4j3})$	0.4123	0.4123	0.0808	0.0946

表 11 空间域融合后的结果 (复杂状态实验 2)
Table 11 Results of fusion in the spatial domain (Complex condition 2)

	h_1	h_2	h_3	U
$m(h_{j1})$	0.6109	0.0162	0.3728	0
$m(h_{j2})$	0.0164	0.7998	0.1837	0
$m(h_{j3})$	0.0001	0.9973	0.0026	0

3.6 与基于贝叶斯方法的行为分析与推理决策的比较

贝叶斯方法由英国学者贝叶斯 (T. Bayes 1701~1761) 提出的贝叶斯公式发展而来. 贝叶斯公式在形式上是对条件概率的定义和全概率公式的一个简单的归纳和推理. 设 A 为样本空间 Ω 中的一个事件, B_1, B_2, \dots, B_n (n 为有限或无穷) 是样本空

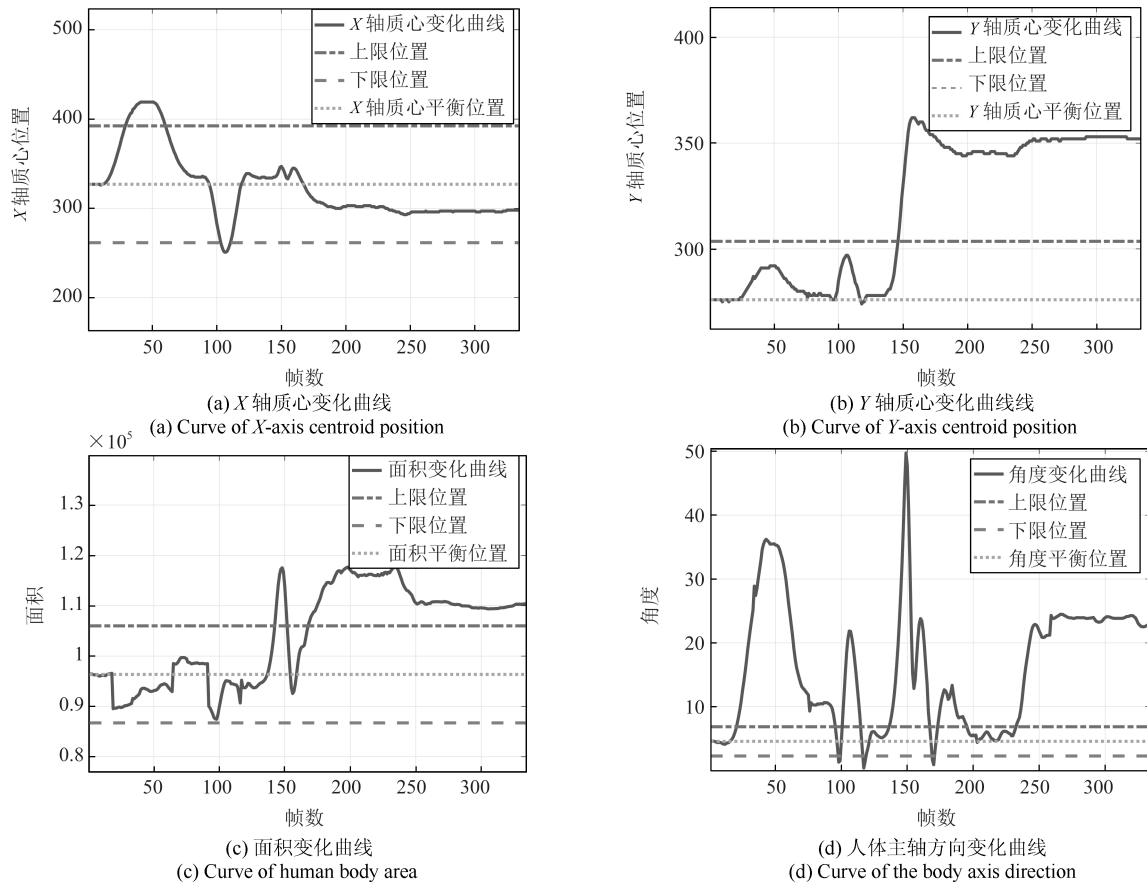


图 15 复杂状态 2 的特征变化曲线

Fig. 15 Characteristic curves of complex state 2 direction

间 Ω 中的一个完备事件群, 且有 $P(B_i) > 0 (i = 1, 2, \dots, n)$, $P(A) > 0$, 则事件 A 的全概率公式可写为式 (18).

$$P(A) = P\left(\sum_{i=1}^n AB_i\right) = \sum_{i=1}^n P(A|B_i)P(B_i) \quad (18)$$

在全概率公式的条件下, 按照条件概率, 贝叶斯公式可以表示为式 (19)

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{P(A)} = \frac{P(A|B_i)P(B_i)}{\sum_{i=1}^n P(A|B_i)P(B_i)} \quad (19)$$

3.6.1 对比实验 1

在这个实验中, 我们使用第 3.4 节, 也就是复杂状态实验 1 的实验数据.

先验概率是指缺乏某个事实的情况下描述一个变量. 在这里的先验概率是指在不知道视频中学生 X, Y, S, θ 变化量的情况下, 而根据专家经验估计学生出现左顾右盼、埋头和端坐的概率. 通过观察视频中的学生状态并和多位有经验的教师分析讨论, 我

们设定: $P(h_1^*) = 0.2, P(h_2^*) = 0.1, P(h_3^*) = 0.7$.

似然函数是指在已知某些观测所得到的结果时, 对有关事物性质的参数进行估计, 即在给定输出 h 时, 关于参数 M 的似然函数 $P(M|h)$, 它 (在数值上) 等于已知观察结果 h 前提下给定参数 M 的发生的条件概率. 例如抛 10 次硬币都为正面朝上的概率, 那硬币正反面对称 (也就是正反面概率均为 0.5 的概率) 的似然程度是多少. 在本文中我们有 4 个证据体 $M (X, Y, \text{面积 } S, \text{角度 } \theta)$, 人体框架中有三个状态 h (左顾右盼、埋头、端坐), 根据理论常识结合多次实验我们发现, 当左顾右盼 (h_1) 时, X 轴质心和角度一定会发生较大的变化, 所以这里我们赋予 $P(X|h_1) = 0.4$ 和 $P(\theta|h_1) = 0.45$ 较大的概率, 而 Y 轴方向上的质心和学生的面积可能不会太大的变化, 所以我们设定 $P(Y|h_1) = 0.1$ 和 $P(S|h_1) = 0.05$. 同样的道理, 我们设定:

$$\begin{aligned} P(X|h_2) &= 0.05, P(Y|h_2) = 0.55 \\ P(S|h_2) &= 0.35, P(\theta|h_2) = 0.05 \\ P(X|h_3) &= 0.25, P(Y|h_3) = 0.25 \\ P(S|h_3) &= 0.25, P(\theta|h_3) = 0.25 \end{aligned}$$

后验概率是指考虑了一个事实之后的条件概率,它通过贝叶斯公式,用先验概率和似然函数计算出来.根据以上的分析,并根据贝叶斯公式(19),那么后验概率为:

$$P(h_1|X) = \frac{P(X|h_1)P(h_1^*)}{\sum_{i=1}^3 P(X|h_i)P(h_i^*)} = 0.3077$$

$$P(h_1|Y) = 0.0800, P(h_1|S) = 0.0455$$

$$P(h_1|\theta) = 0.3333$$

同理得到:

$$P(h_2|X) = 0.0192, P(h_2|Y) = 0.22$$

$$P(h_2|S) = 0.1591, P(h_2|\theta) = 0.0185$$

$$P(h_3|X) = 0.6731, P(h_3|Y) = 0.7000$$

$$P(h_3|S) = 0.7954, P(h_3|\theta) = 0.6482$$

从复杂状态实验 1 的实验数据中,结合式(10)~(14),我们可以得到,在这段时间内, X 轴质心变化的帧数为 36, Y 轴质心变化的帧数为 7,面积变化的帧数为 3,角度变化的帧数为 35.所以 $P(X) = 0.4444, P(Y) = 0.0864, P(S) = 0.0370, P(a) = 0.4321$.

再根据全概率公式,本次实验中左顾右盼的概率为:

$$\begin{aligned} P(h_1) &= \sum_{j=1}^4 P(h_1|M_j)P(M_j) = \\ & \sum_{j=1}^4 P(h_1|M_j) \sum_{i=1}^3 P(M_j|h_i)P(h_i^*) = \\ & \sum_{j=1}^4 \sum_{i=1}^3 P(h_1|M_j)P(M_j|h_i)P(h_i^*) = \\ & 0.2894 \end{aligned}$$

其中, $P(M_j)$ 表示的是 4 个证据体(X 轴质心坐标、 Y 轴质心坐标、面积和角度)发生变化的概率.

同样的方法,我们可以得到埋头的概率为:

$$P(h_2) = \sum_{j=1}^4 P(h_2|M_j)P(M_j) = 0.0415$$

端坐的概率为:

$$P(h_3) = \sum_{j=1}^4 P(h_3|M_j)P(M_j) = 0.6692$$

根据上面得到的概率值,我们可以得到这段时间学生处于端坐状态,可以推测其学习比较认真.

为了证明不同的先验概率对最后决策结果的影响,这里加入另一组先验概率下的决策推断.设定 $P(h_1^*) = 0.3, P(h_2^*) = 0.1, P(h_3^*) = 0.6$. 最终得出的 $P(h_1) = 0.3805, P(h_2) = 0.0534, P(h_3) = 0.5662$. 从结果可以看出不同的先验概率,虽然会影响最终的概率值,但不会影响最终的决策结果.

3.6.2 对比实验 2

在这个实验中,我们使用第 3.5 节也就是复杂状态实验 2 的实验数据.根据专家经验估计并通过观察视频中的学生状态并和多位有经验的教师分析讨论,我们设定先验概率为 $P(h_1^*) = 0.2, P(h_2^*) = 0.65, P(h_3^*) = 0.15$. 由于在抛硬币时,不管第几次抛,何时抛,硬币正反面对称(也就是正反面概率均为 0.5 的概率)的似然程度是不变的,所以本次实验的似然函数仍采用对比实验 1 中的值.

有了先验概率和似然函数,根据贝叶斯公式,那么后验概率为:

$$P(h_1|X) = 0.5333, P(h_1|Y) = 0.0482$$

$$P(h_1|S) = 0.0364, P(h_1|A) = 0.5625$$

$$P(h_2|X) = 0.2167, P(h_2|Y) = 0.8614$$

$$P(h_2|S) = 0.8273, P(h_2|A) = 0.2031$$

$$P(h_3|X) = 0.2500, P(h_3|Y) = 0.0904$$

$$P(h_3|S) = 0.1364, P(h_3|A) = 0.2344$$

通过实验数据我们得到:

$$P(X) = 0.0583, P(Y) = 0.2883$$

$$P(S) = 0.2699, P(A) = 0.3834$$

最后根据全概率公式,得到:

$$P(h_1) = 0.2705, P(h_2) = 0.5622$$

$$P(h_3) = 0.1673$$

在表 12 中我们比较了使用 D-S 证据理论和贝叶斯方法得到的结果,可以发现和使用贝叶斯方法相比,使用 D-S 理论可以降低推理过程中的不确定性,同时使用 D-S 方法得到的结果也与我们的认知更为相符.

表 12 D-S 推理方法和贝叶斯推理方法的对比
Table 12 Comparison of D-S inference and the Bayesian inference methods

对比实验	理论方法	h_1	h_2	h_3	U
对比实验 1	D-S 理论	0.0001	0	0.9999	0
	贝叶斯方法	0.2894	0.0415	0.6692	-
对比实验 2	D-S 理论	0.0001	0.9863	0.0136	0
	贝叶斯方法	0.2705	0.5622	0.1673	-

从表 12 中的数据我们可以看出, D-S 理论方法比贝叶斯方法具有更高的决策支持. 在对比实验 1 中, D-S 理论得出的结果相对于贝叶斯理论方法提高了 $\frac{0.9999-0.6692}{0.6692} \times 100\% = 49.42\%$ 的决策支持; 而在对比实验 2 中, 相对于贝叶斯方法, D-S 理论更是提高了 $\frac{0.9863-0.5622}{0.5622} \times 100\% = 75.44\%$ 的决策支持, 贝叶斯方法过多地依赖于先验概率和经验常识, 使用 D-S 理论可以更好地处理得到的人体信息, 并完成决策推理工作. 与贝叶斯方法的对比实验表明, 使用 D-S 理论可以大幅度降低决策中的不确定性, 同时得到的决策结果也与我们的认知更加符合.

4 结论

为了解决在线课程授课过程中, 缺乏对于学生学习情况的跟踪与教学效果评估问题, 本文依据视频信息对学生注意力进行建模, 并提出了一种评判学生听课专心程度的行为自动分析算法. 该算法能够有效跟踪学生的学习状态, 提取学生的行为特征参数, 并对这些参数进行 D-S 融合判决, 以获得学生的听课专注度. 经过多次实验的结果表明, 本文采用的方法能够有效评判学生在授课期间的专心程度, 对每一种行为的识别率可以达到 95% 以上. 在融合决策中, 与贝叶斯推理方法相比, 采用 D-S 融合评判的准确度更高.

本文还存在着分类的种类较为简单, 不能完全概括网络在线授课中学生的学习状态的不足. 对此, 可以扩充人体行为的分类方式, 尝试加入模糊隶属度函数, 在多种分类的情况下, 对人体的行为进行分析和判断, 得到更为完备的人体辨识框架. 这也是本研究后续的工作方向.

References

- 1 Wong K, Patzelt M, Poulette B, Hathaway R. Scenario-based learning in a MOOC specialization capstone on software product management. In: Proceedings of the IEEE/ACM 39th International Conference on Software Engineering Companion (ICSE-C). Buenos Aires, Argentina: IEEE, 2017. 317–318
- 2 Staubitz T, Willems C, Hagedorn C, Meinel C. The gamification of a MOOC platform. In: Proceedings of the 2017 IEEE Global Engineering Education Conference (EDUCON). Athens, Greece: IEEE, 2017. 383–392
- 3 Lei C U, Yeung Y C A, Kwok T T O, Lau R, Ang A. Leveraging videos and forums for small-class learning experience in a MOOC environment. In: Proceedings of the 2016 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE). Bangkok, Thailand: IEEE, 2016. 409–411
- 4 Luo H, Millet A, Alley R, Zuo M. Dealing with ethical issues in MOOC design and delivery: a case study. In: Proceedings of the 2016 Blended Learning: Aligning Theory with Practices. ICBL 2016. Lecture Notes in Computer Science, vol. 9757. Cham, Switzerland: Springer, 2016. 128–138
- 5 Kearney R C, Premaraj S, Smith B M, Olson G W, Williamson A E, Romanos G. Massive open online courses in dental education: two viewpoints: viewpoint 1: massive open online courses offer transformative technology for dental education and viewpoint 2: massive open online courses are not ready for primetime. *Journal of Dental Education*, 2016, **80**(2): 121–127
- 6 Brinton C G, Buccapatnam S, Wong F M F, Chiang M, Poor H V. Social learning networks: efficiency optimization for MOOC forums. In: Proceedings of the IEEE INFOCOM 2016 — The 35th Annual IEEE International Conference on Computer Communications. San Francisco, USA: IEEE, 2016. 1–9
- 7 Jaouedi N, Boujnah N, Htiwich O, Bouhlel M S. Human action recognition to human behavior analysis. In: Proceedings of the 7th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT). Hammamet, Tunisia: IEEE, 2016. 263–266
- 8 Xiao B, Georgiou P, Baucom B, Narayanan S S. Head motion modeling for human behavior analysis in dyadic interaction. *IEEE Transactions on Multimedia*, 2015, **17**(7): 1107–1119
- 9 Tsai H C, Chuang C H, Tseng S P, Wang J F. The optical flow-based analysis of human behavior-specific system. In: Proceedings of the 1st International Conference on Orange Technologies (ICOT). Tainan, China: IEEE, 2013. 214–218
- 10 Batchuluun G, Kim J H, Hong H G, Kang J K, Park K R. Fuzzy system based human behavior recognition by combining behavior prediction and recognition. *Expert Systems with Applications*, 2017, **81**: 108–133
- 11 Alonso J B, Cabrera J, Travieso C M, López-de-Ipiña K. First approach to continuous tracking of emotional temperature. In: Proceedings of the 4th International Work Conference on Bioinspired Intelligence (IWobi). San Sebastian, Spain: IEEE, 2015. 177–184
- 12 Hwang K A, Yang C H. Attentiveness assessment in learning based on fuzzy logic analysis. *Expert Systems with Applications*, 2009, **36**(3): 6261–6265
- 13 Durucan E, Ebrahimi T. Moving object detection between multiple and color images. In: Proceedings of the 2013 IEEE Conference on Advanced Video and Signal Based Surveillance. Miami, FL, USA: IEEE, 2003. 243–251
- 14 Hu M J, Liu Z, Zhang J Y, Zhang G J. Robust object tracking via multi-cue fusion. *Signal Processing*, 2017, **139**: 86–95
- 15 Zhang P, Zhuo T, Huang W, Chen K L, Kankanhalli M. Online object tracking based on CNN with spatial-temporal saliency guided sampling. *Neurocomputing*, 2017, **257**: 115–127

- 16 Bilge Y C, Kaya F, Cinbis N İ, Celikkan U, Sever H. Anomaly detection using improved background subtraction. In: Proceedings of the 25th Signal Processing and Communications Applications Conference (SIU). Antalya, Turkey: IEEE, 2017. 1–4
- 17 Elyounsi A, Tlijani H, Bouhlef M S. Shape detection by mathematical morphology techniques for radar target classification. In: Proceedings of the 17th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering (STA). Sousse, Tunisia: IEEE, 2016. 352–356
- 18 Shi X S, Huang Y J, Liu Y G. Text on Oracle rubbing segmentation method based on connected domain. In: Proceedings of the 2016 IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference. Xi'an, China: IEEE, 2016. 414–418
- 19 Shi H B, Yang S Y, Cao Z L, Pan W, Li W H. An improved evidence combination method of D-S theory. In: Proceedings of the 2016 International Symposium on Computer, Consumer and Control. Xi'an, China: IEEE, 2016. 319–322
- 20 Chu Q, Ouyang W L, Li H S, Wang X G, Liu B, Yu N H. Online multi-object tracking using CNN-based single object tracker with spatial-temporal attention mechanism. In: Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE. 2017. 4846–4855



戴亚平 北京理工大学自动化学院教授。主要研究方向为图像特征提取与识别, 多传感器数据融合与决策诊断技术, 人工智能与专家系统。

E-mail: daiyaping@bit.edu.cn

(**DAI Ya-Ping** Professor at the School of Automation, Beijing Institute of Technology, China. Her research

interest covers image feature extraction and recognition, multi-sensor data fusion and decision diagnosis technology, artificial intelligence and expert system.)



杨方方 北京理工大学自动化学院硕士研究生。2016 年获得东北大学学士学位。主要研究方向为计算机视觉与机器学习。

E-mail: neuqyangfangfang@163.com

(**YANG Fang-Fang** Master student at School of Automation, Beijing Institute of Technology, China. She received her bachelor degree from School of Control Engineering, Northeastern University, China. Her research interest covers computer vision and machine learning.)

control Engineering, Northeastern University, China. Her research interest covers computer vision and machine learning.)



赵翰奕 华北计算机系统工程研究所(信息产业部电子第六研究所) 研究工程师。2017 年获得北京理工大学自动化学院硕士学位。主要研究方向为图像处理 and 深度学习。

E-mail: zhao_angus@163.com

(**ZHAO Han-Yi** Research engineer at The 6th Research Institute of China

Electronics Corporation. He received his master degree from School of Automation, Beijing Institute of Technology, China in 2017. His research interest covers image processing and deep learning.)



贾之阳 北京理工大学自动化学院助理教授。2010 年和 2013 年分别获得西北工业大学电气工程系学士学位和北京工业大学控制科学与工程系硕士学位。2017 年获得美国康涅狄格大学电气工程系博士学位。主要研究方向为智能制造, 生产系统建模, 性能分析, 能源高效生产管理。本文通信作者。

E-mail: zhiyang.jia@bit.edu.cn

(**JIA Zhi-Yang** Assistant professor at the School of Automation, Beijing Institute of Technology. He received his bachelor degree and master degree from Department of Electrical Engineering, Northwestern Polytechnical University, and Department of Control Science and Engineering, Beijing University of Technology, China in 2010 and 2013. He received his Ph. D. degree from Department of Electrical and Computer Engineering, University of Connecticut, USA in 2017. His research interest covers smart manufacturing, modeling, analysis and control of production systems. Corresponding author of this paper.)



广田熏 北京理工大学自动化学院“外专千人”特聘教授, 1976 年和 1979 年分别获得东京工业大学电子工程专业硕士学位、博士学位。主要研究方向为图像模式识别, 智能机器人, 模糊控制, 人工智能及工业应用。

E-mail: hirota@bit.edu.cn

(**HIROTA Kaoru** Special professor of The Recruitment Program of Global Experts at the

School of Automation, Beijing Institute of Technology, China. He received his master degree and Ph. D. degree from Electric Engineering, Tokyo Institute of Technology, Japan in 1976 and 1979. His research interest covers image pattern recognition, intelligent robotics, fuzzy control, artificial intelligence, and industrial applications of these topics.)