

基于神经网络的学习状态检测

郑茜元, 郑虹, 侯秀萍

(长春工业大学计算机科学与工程学院, 吉林 长春 130000)

✉17643093205@163.com; zhenghong@ccut.edu.cn; houxiuping@ccut.edu.cn



摘要: 对在线学习者注意力状态检测的方法大多基于眼睛闭合频率、头部偏转等特征, 此类方法能够应对大多数情况, 但针对学习者正视屏幕且视线落点处于屏幕上时出现的发呆、分神状态无法作出检测。针对此问题, 提出了一种基于RNN的眼动分析算法RNN-EMA(RNN-Eye Movement Analysis), 该算法通过对序列眼动向量分析, 预测学生学习行为, 完成当前学习状态检测。实验表明, RNN-EMA算法能够对学习状态作出有效检测, 且对比同类方法效果有所提升。

关键词: 在线学习; 循环神经网络; 眼动分析; 注意力检测

中图分类号: TP311.5 **文献标识码:** A

Detection of the Learning State Based on Neural Network

ZHENG Qianyuan, ZHENG Hong, HOU Xiuping

(School of Computer Science and Engineering, Changchun University of Technology, Changchun 130000, China)

✉17643093205@163.com; zhenghong@ccut.edu.cn; houxiuping@ccut.edu.cn

Abstract: Online learners' attention states are mostly detected through eye closure frequency, head rotation and other action features. These methods can cope with most situations, but cannot detect the absent-minded and distracted state when the learner is facing the screen and the sight point is on the screen. To solve this problem, the paper proposes an RNN-EMA (RNN-Eye Movement Analysis) algorithm based on RNN. The algorithm predicts the learning behavior of students through sequential eye movement vector analysis, and conducts the current learning state detection. Experiments show that the RNN-EMA algorithm can effectively detect the learning state, and the accuracy is improved compared with other methods of the same kind.

Keywords: online learning; RNN; eye movement analysis; attention detection

1 引言(Introduction)

网络在线学习不受时间与地域的限制, 在现代教育中得到广泛使用。学习者在接受知识的过程中, 由于缺乏有效地监督, 导致其学习效果无法得到保证, 因此, 需要对学习者的注意力状态进行有效监督, 以保证在线学习者的学习质量。

对于学习状态的研究包括心境状态、注意力状态和动机状态等。研究的方法大多基于生物特征计算^[1,2]、学习行为分析^[3]、面部特征分析^[4,5]等方面。吴沧海^[6]等人对学习者的表情、眼睛闭合特征、体姿进行分析, 通过情感计算得出学习状态。此类方法在处理学习者正视屏幕时出现的分神状态无法进行有效检测。

针对上述问题, 本文对学习者的眼动行为进行分析, 实现对学习者学习状态的判断。相对于传统学习状态检测方法, 能更深入地对学习过程中的细节特征进行分析, 得到更准确的判断。

2 眼动分析研究(Eye movement analysis research)

人类对信息的获取很大程度上依赖于视觉, 故人类的视觉与学习过程中的注意力状态有很大的相关性。近年来, 对眼动行为的分析逐渐成为热门研究课题。借此灵感, 一些学者通过对眼动行为分析, 估计用户的注意力状态。王凤娇^[7]等人提出了一种CMVA分类视觉注意模型, 该方法将自顶向下和自底向上注意相融合, 通过眼动数据对模型进行训练, 实

现了人在观察图片时显著性区域预测，在自己采集的数据集上进行实验，实验结果表明，与其他算法相比，该方法具有较高的预测准确率。但此方法只实现了对人们观看图片时的显著性区域预测，并未考虑注意力集中程度的问题。江波^[8]等人在网络在线测评的环境下，利用Tobii EyeX眼动仪记录了答题者在答题过程中的眼动行为，以此探索了答题者在答题过程中眼动行为轨迹与视觉注意模式之间的联系，但此类方法对设备仪器有一定的要求，且入侵式检测易给学生带来压力。王晓明^[9]等人为实现对阅读者的注视点进行预测，以双向长短期记忆—条件随机场神经网络为基础，通过用户在阅读过程中的眼动数据对模型进行训练，得到视觉注意模型，实现了对阅读者视线焦点的预测，但该方法同样未实现对注意力状态是否集中有一个明确判断。Yi J^[10]等人基于普通摄像头对在线学习者的面部图像进行采集，提取图像中的眼动特征，以连续的眼动向量为输入，对HMM模型进行训练，实现了对学习者注意力状态的检测。

然而HMM模型具有一定的局限性，HMM的独立性假设忽略了帧间的相关性，齐次马尔科夫假设使得任意时刻的隐藏状态只依赖于它前一个隐藏状态，降低了模型对信息的表示能力。为克服此问题，本文提出了一种基于RNN的眼动分析算法RNN-EMA。引入处理序列问题的循环神经网络(RNN)，考虑每个隐藏状态与之前隐藏状态和当前时刻输入的关联关系，采用神经元的规模和层数对数据进行拟合得出潜在结构。通过实验验证，该算法在眼动行为分类方面表现出良好的性能，分类准确度有明显提升。

3 RNN-EMA眼动分析算法(RNN-EMA eye movement analysis algorithm)

3.1 眼动行为描述

据研究表明，人类在获取并加工信息时，眼睛并非快速扫过所观看事物，而是视线焦点滞留在信息的某一具体位置对信息进行充分的加工处理，处理完毕后，通过“眼跳”到下一个位置。根据眼睛的运动方式，可用两个术语“固着”(Fixation)、“扫动”(Saccade)来描述一系列眼动行为^[10]。其定义如下：

- ① “固着”状态：指当视线处于相对静止时的状态。
- ② “扫动”状态：指当从一个固着状态到另一个固着状态，即发生眼跳，称之为“扫动”。

本文对眼动行为进行分类的依据是眼动向量，利用深度学习的方法对眼动向量进行特征提取，实现眼动分类，由此得出当前学习者学习状态。通过“固着”“扫动”两个状态对眼动行为进行划分，形成眼动向量。通过对眼动行为的观察，将眼睛在认知过程中的动作分为阅读、搜索、分心三种模式，对序列眼动向量进行分析，将其分类到三种眼动模式中，实现对学习者注意力状态评估。

3.2 眼动向量定义

眼动向量的定义以当前学习者的眼角位置为基准，分析采集到的每一帧图像中虹膜中心位置相对于眼角点的偏移量，以此判断当前眼动行为。

定义：眼动向量 $M = (X, Y, T)$ ，其中， X 表示当前固着状态较前一固着状态在水平方向的位移， Y 表示当前固着状态较前一固着状态在垂直方向的位移， T 表示此次眼动行为的持续时间。用 M_k 表示捕捉到的第 K 个眼动向量， $M_k = (X_k, Y_k, T_k)$ ，通过以下公式求取 M_k 的各个分量值：

$$X_k^0 = \sum_{f_i \in S_k} x_i / n, i = 1, 2, \dots, n \tag{1}$$

$$Y_k^0 = \sum_{f_i \in S_k} y_i / n, i = 1, 2, \dots, n$$

$$\begin{aligned} X_k &= X_k^0 - X_{k-1}^0 \\ Y_k &= Y_k^0 - Y_{k-1}^0 \end{aligned} \tag{2}$$

$$T_k = \sum_{f_i \in S_k} t_i, i = 1, 2, \dots, n \tag{3}$$

其中， t 表示从采集到的图像中计算分析得到的视线状态，每一帧图像的视线状态由三元组 (x, y, t) 表示，其中 x 、 y 分别表示虹膜中心相对于眼角点在水平方向和垂直方向上的距离变化， t 表示采样每帧图像所需的时间。每一个固着状态表现为若干连续且视线状态相似的帧，用 S_k 表示一个固着状态中所有视线记录的集合。 S_k 集合中记录的条数用 n 表示， T_k 表示这一固着状态持续的时间。

3.3 眼动分类模型

眼动分析输入一个序列的眼动向量，网络最终输出这个眼动序列属于哪种眼动模式，采用多输入单输出的结构，在输入方面，输入的是通过提取得到的眼动向量，具体模型结构，如图1所示。

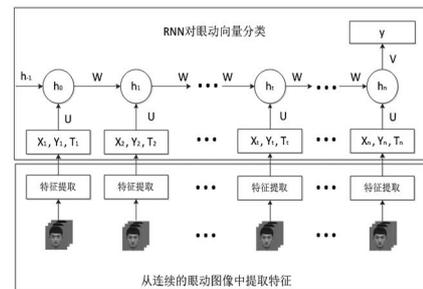


图1 RNN-EMA眼动分析模型

Fig.1 RNN-EMA eye movement analysis model

由图1可知，对视频流中的连续帧先进行特征提取，得到一系列的眼动向量 (X_i, Y_i, T_i) ，这样减少了网络对基本特征的寻找，着重于眼动特征分析。将连续的眼动向量作为循环神经网络的输入，对网络进行训练。 h_t 为第 t 步的隐藏状态， h_1 为初始状态， U 、 V 、 W 为循环神经网络中的参数，在训练过程中共享， Y 为最终的输出结果，结果可取0、1、2三种值，分别代表“阅读”“搜索”“分心”三种状态。其中阅读和搜索两种状态为注意力集中状态，分心状态为注意力不集中状态。

4 眼动分析实验(Eye movement analysis experiment)

4.1 数据集采集

由于此类研究没有公开的数据集，采用收集的数据进行实验。为收集到不同状态下的眼动特征，制定一些合理的任务，使志愿者产生不同的眼动行为，在完成任过程中收集

志愿者的眼动特征，计算眼动向量。“阅读”“搜索”“分心”三种状态的任务设计如下：

阅读状态：在屏幕上呈现一段文字或一个物体，要求志愿者去认真地阅读这段文字或者观察物体，为避免学习者提前完成任务，我们会设置足够的阅读量，尽量避免错误数据的录入。

搜索状态：给定文章，要求志愿者去寻找指定的词。

分心状态：为了使志愿者的注意力尽力集中到视觉以外的其他感官上，在屏幕上显示空白图片，同时给志愿者播放一段音乐，或者要求志愿者去回忆一段往事。

对收集到的数据进行清洗处理后，最终收集到1065条眼动向量，其中阅读状态340条，搜索状态482条，空闲状态243条。将这些眼动数据按7:3的比例分为训练集和测试集。

4.2 噪声处理

本文以虹膜中心的运动为基础提取眼动向量，虽然极力的在每一个步骤都去做好优化以提升虹膜中心定位准确率，但不可否认的是依然会存在一定的误差。为避免误差点致使整体眼动行为分析的准确率降低，采用欧氏距离度量两帧图像中虹膜中心相似度。具体公式如下所示：

$$\sqrt{(x_0 - x_1)^2 + (y_0 - y_1)^2} < d \tag{4}$$

其中， x_0 、 y_0 表示上一帧图像中虹膜中心距离眼角点的距离， x_1 、 y_1 表示当前帧虹膜中心距离眼角点的距离。当两次变化的欧式距离小于设定阈值时，则认为是一次眼动行为；大于阈值时，判定为二次眼动。

此方法能够将足够相似的视线状态聚集在同一次眼动行为中，当提取到新的视线状态时，通过阈值的判断，决定是否将其归类为当前眼动行为，通过该方法，能够对噪声数据进行有效处理，增强算法的稳定性，提升眼动分类准确率。

4.3 防止过拟合策略

由于数据量较少，为预防出现过拟合问题，本文采取以下策略：(1)设计的模型较简单，模型复杂度较低；(2)输入之前对图像进行处理，提取出较少且关键的特征，在对实验效果影响有限的情况下有效防止了过拟合；(3)采用Dropout操作，随机丢弃50%的结点，防止过拟合现象。

4.4 结果及分析

本文实验均在Intel(R)Core(TM) i5-4590S，CPU主频3.00GHz，8GB内存的Windows操作系统，Python集成环境anaconda3环境下运行。摄像头采用电脑自带摄像头。

RNN-EMA算法实验基于TensorFlow深度学习框架，以6组眼动向量为一个序列输入到RNN的输入层中，设置循环神经网络的隐藏层单元数为10，学习率为0.0001，目标函数使用softmax交叉熵函数，使用AdamOptimizer方法对循环神经网络进行优化。在训练集上进行30000次迭代训练，选取其中最好的训练模型保存，当训练到第6000次的时候损失值趋于收敛。

为验证RNN-EMA眼动分析算法的性能，我们采用Python3.6语言还原了文献[10]中的方法，同样以6组的数据为一个序列输入模型中，分别训练三个HMM，最终将测试集中

的眼动序列分别输入三个HMM中，计算每个HMM的发生概率，将这段眼动序列归类为发生概率最大的模型所对应的眼动状态。

使用准确率(Accuracy)和混淆矩阵(Confusion Matrix)两种分类器评估方法对模型进行评估。最终准确率，详见表1。

表1 眼动分析准确率对比

Tab.1 Comparison of accuracy of eye movement analysis

| 评估方法 | 文献[10]算法 | 本文算法 |
|------|----------|--------|
| 准确率 | 92.1% | 97.37% |

由表1可以看出，与文献[10]中提出的算法相比，本文提出的算法有较好的效果，在准确率上较对比方法提高了5.27%。这是由于，相对于对比文献中的方法，本文方法考虑了当前隐藏状态与之前隐藏状态和当前时刻输入的相关性，提升了模型对数据的表达能力，从而分类准确率得到了提升。

为了更加直观地看到模型对每个类别的分类准确率，深入分析两种方法在分类准确度，以及性能上的差异，我们展示出每个模型的混淆矩阵，如图2所示。

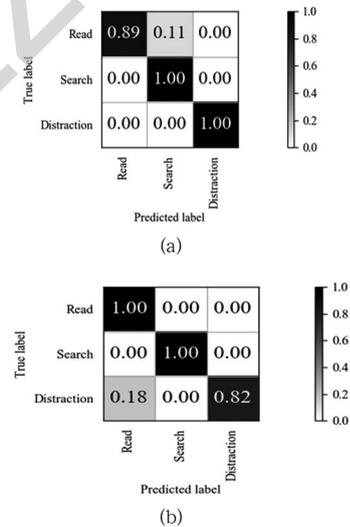


图2 算法对比混淆矩阵

Fig.2 Algorithm contrast confusion matrix

如图2所示，图2(a)为RNN-EMA眼动分析方法的混淆矩阵，图2(b)为文献[10]中方法的混淆矩阵，分析可得两种算法对“搜索”状态都能达到100%的分类准确率，效果良好，这可能是由于搜索状态的特征相对于其他两种状态有较大的差异性，模型更容易识别。两种算法性能的对“阅读”和“分心”两种状态的分类，文献[10]中的方法对“分心”状态的分类准确率为82%，有18%的错误率将“分心”状态误判为“阅读”状态，而本文提出的算法对“分心”状态的分类准确率达到100%，对“阅读”状态的分类准确率为89%，有11%的误差将“阅读”状态误判为“搜索”状态。由于本文的着重点是注意力状态检测，所以眼动检测更侧重于检测出“分心”状态，从这一点来考虑，本文所提出

(下转第5页)