Students' Attention Assessment in eLearning based on Machine Learning

Qingshan Deng[1], Zhili Wu[1]

[1]School of Software & Internet of Things Engineering, Jiangxi University of Finance and Economics, Nanchang 330013, Jiangxi, China

Corresponding author: Qingshan Deng (dqs@jxufe.edu.cn)

Abstract. Multimedia-based Electronic learning (eLearning) is an effective method of knowledge transfer. It provides the opportunity that students can use the videos or other materials at any time after they are delivered. Multimedia applications provide convenience, but there also exist challenges. One of the challenges is to measure and assess students' attention when they are studying online. This paper presents a framework based on machine learning methods for the measurement of students' attention. The framework employs a Gabor wavelet to extract the eye state features and train the model using support vector machines (SVM) to complete automatic classification on students' eye states. Experiments over thousands of facial photos show that the proposed system reaches a good performance, which has a significant value in real applications.

1. Introduction

ELearning is a way to teach and learn online in a cyber virtual classroom through the computer or mobile phone with wireless network. Recently, eLearning has developed very rapidly, since eLearning can promote learning in a variety of ways. Students can learn at any time, any place and self-paced under eLearning environment [1,2]. In addition, the development of information communication technology (ICT) and multimedia-based technologies has created potential benefits for students [3]. Multimedia and ICT also provide opportunities and conditions for monitoring and analysing students' learning behaviour and predicting its trends and patterns. Such analysis can help teachers design a new and effective teaching method and present teaching content in a better way [4]. The combination of advances in the smart device technology and the intelligent assessment systems has facilitated the use of multimedia-based educational resources without restrictions on time and place. These educational materials are delivered in the form of data and compute-intensive multimedia-enabled learning objects [5]. Academics and industry are developing online learning systems that turn society into a ubiquitous educational institution [6].

At present, the assessment of the effect of eLearning is usually based on active feedback from students or teachers. However, some studies state that a large amount of data can be well organized to construct semantics [7, 8]. Therefore, this paper presents a qualitative assessment method for visual attention of students. The purpose of this research is to analyse the students' video stream during learning and extracts the visual attention features to verify the attention problems of students.

This paper presents a machine learning-based assessment framework for analysing student's attention in an eLearning environment. In this article, we mainly focus on students' eye states classification and propose an automatic assessment model based on machine learning. The method uses Gabor to extract the eye states features and uses the support vector machines (SVM) algorithm as the classifier. Experiments show that this model achieves good performance both in robustness and correctness, which has significant value in real applications. In addition, the framework including the automatic face detection and feature points detection based on current machine learning models is proposed. These models are labour-saving and provide more precise detection than manfully labelling. In summary, the following are the main contributions of this paper.

1. We propose a machine learning framework for students' visual attention assessment. The framework was able to measure the students' attentiveness based on automatic detection of the faces especially facial points and eye states.
2. A machine learning-based model is presented for eye state classification. The model is based on Gabor feature, which is very similar to the visual stimuli response of simple cells in the human visual system, and support vector machines (SVM). The model can obtain the accuracy to 93.1% which outperforms other current machine learning models.

2. Visual Attention Assessment

We proposed a framework to measure the student's visual attention automatically. This framework starts working by extracting frames from the video stream of the student's webcam. Each frame is automatically analysed by searching the faces including facial points and eye states. Eyes' states are detected automatically to decide whether they are opened or closed. This automatic detection is based on machine learning algorithms and predictors [16]. Fig. 1 shows the schematic diagram of the proposed framework, which includes the following steps.

1. **Face Detection.** Firstly, we should determine whether there is a face or not in an image or video sequence. The score is calculated for each frame according to formula (1).

$$Score_{face}(f) = y\begin{cases} 0, & if\ no\ face \\ \sum_{i=1}^{n} f_i\,, & on\ each\ face \end{cases} \quad (1)$$

2. **Facial Point Detection.** For further analysis, we generate the 68 facial points by current machine learning model [16]. The 68 facial points cover the main parts of the face including eyes, mouth, nose and eyebrows, which are famous on the face detection.
3. **Eye Detection.** After the facial points are obtained, we detect the eye area-related points and extract the eye images. The score is calculated as,

$$Score_{eye}(f) = x\begin{cases} 0, & if\ no\ eye \\ \sum_{i=1}^{n} e_i\,, & on\ each\ face \end{cases} \quad (2)$$

4. **Eye state classification.** After obtaining the eye photos. The score is calculated for each eye,

$$eyeScore(e) = \sum_{i=1}^{n} eye_i \begin{cases} 0, & if\ eye\ is\ closed \\ 1, & if\ eye\ is\ open \end{cases} \quad (3)$$

5. **Attention scoring.** The students' attentiveness based on the face detection and the eye state can be divided into 5 levels: Away, Sleepy, Mind wandering, Satisfactory and Attentive according to the attention scores.

$AtentionScore =$

$$eyeScore(e) + Score_{eye}(f) + Score_{face}(f) \quad (4)$$

The facial points are generated by the current machine learning model as shown in the middle in Fig. 1. We can utilize this model in the first three steps. However, the eye state classification is another difficult problem in this system. To solve this problem, we proposed a machine learning model consisting of feature extraction and classification algorithms for the eye state classification, which would be introduced in the next section.

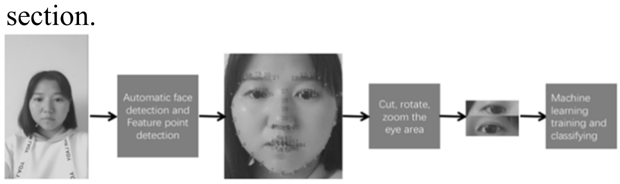

Figure 1. Student's visual attention assessment framework based on machine learning models.

3. Eye State Classification

For eye state classification, we propose to use machine learning approach to automatically give the eye score. Machine learning algorithms are mainly divided into two parts, feature detection and classifier algorithms. This article compares existing several feature extraction algorithms and classifier algorithms. Feature extraction algorithms include principal components analysis (PCA) and Gabor features. The classification algorithms include k-nearest neighbours algorithm (KNN), naive Bayes classifiers (NBC) and support vector machines (SVM). The algorithms are described in detail below.

3.1 Feature Extraction Algorithms

PCA is a technique for analysing and simplifying the data sets. PCA is often used to reduce the dimensionality of a data set while maintaining the largest contribution to the variance in the data set. This is done by retaining the low-order principal components and ignoring the higher-order principal components. These low-level components can often retain the most important aspects of the data. PCA was invented by Carl Pearson in 1901 [9] to analyse data and establish mathematical models. The method is mainly utilizing the feature decomposition of the covariance matrix [10] to obtain the main components of the data (ie, feature vectors) and their weights. PCA provides an effective way to reduce the data dimension and is particularly useful when analysing complex data, such as face recognition.

The Gabor transform belongs to the windowed Fourier transform. The Gabor function can extract relevant features in different scales and directions in the frequency domain [11]. The Gabor wavelet is very similar to the visual stimuli response of simple cells in the human visual system. It performs well in extracting the target's local space and frequency domain information. Although the Gabor wavelet itself does not constitute an orthogonal basis, it can form a tight frame under certain parameters. Gabor wavelet is sensitive to the edge of the image, can provide good direction selection and scale selection characteristics, and is insensitive to changes in light. It can provide good adaptability to changes in light. Gabor has excellent spatial locality and directional selectivity and can capture spatial frequencies (scales) and local structural features in multiple directions in a local region of the image. Gabor decomposition can be seen as a microscope which is sensitive for directions and scales.

Mathematically, a 2D isotropic Gabor filter is the product of a 2D Gaussian and a complex exponential function. The general expression can be expressed as,

$$g_{\theta,\gamma,\sigma}(x,y) = \exp\left(\frac{-(x^2+y^2)}{\sigma^2}\right)\exp\left(\frac{j\pi}{\lambda}(xcos\theta + ysin\theta)\right) \qquad (5)$$

The parameter θ represents the orientation, λ is the wavelength, and σ indicates scale at orthogonal direction. However, with this set of parameters, the Gabor filter does not scale uniformly as the parameter σ changes. It is better to use a parameter $\theta = \lambda/\sigma$ to replace λ so that a change in σ corresponds to a true scale change in the Gabor filter. Also, it is convenient to apply a 90^o counter clockwise rotation to (5), such that θ expresses the normal direction to the Gabor wavelet edges. Therefore, the Gabor filter can be alternatively defined as follows,

$$g_{\theta,\gamma,\sigma}(x,y) = \exp\left(\frac{-(x^2+y^2)}{\sigma^2}\right)\exp\left(\frac{j\pi}{\gamma\sigma}(xcos\theta - ysin\theta)\right) \qquad (6)$$

By selectively changing each of the parameters of the Gabor filter, one can tune the filter to a specific pattern arising in the image. Some examples of Gabor filter with different parameters (γ, θ, σ) are illustrated in Fig. 2.

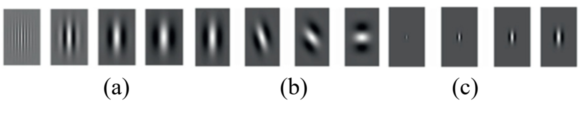

Figure 2. Examples of Gabor filters: Each example shows the real part of Gabor filter for different parameters, (a) $\gamma = \left\{\frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \frac{7}{2}\right\}$; (b) $\theta = \left\{0, \frac{\pi}{6}, \frac{\pi}{3}, \frac{\pi}{2}\right\}$; (c) $\sigma = \{4, 8, 12, 16\}$.

3.2 Classification Algorithms

KNN is a case-based learning, or local approximation and lazy learning after deferring all calculations to classification [12]. Whether it is classification or regression, the measurement of the neighbours' weights is very useful, so that the weights of the closer neighbours are greater than the weights of the neighbours which is far away. For example, a common weighting scheme is to assign a weight of 1/d to each neighbour, where d is the distance to the neighbour.

NBC is a series of probabilistic classifiers based on Bayes theorem using strong (simple) independence between hypotheses [13]. Naive Bayes classifiers are highly scalable and therefore require parameters whose characteristics are linear with the classifier. For some types of probability models, a very good classification effect can be obtained in the sample set of supervised learning. In many practical applications, the NBC's parameter estimation uses the maximum likelihood estimation method. Maximum likelihood training can be done by evaluating a closed-form expression without the need for time-consuming iterative approximations used by many other types of classifiers.

SVM is a machine learning algorithm based on statistical learning theory with minimal structural risk. More formally, a support vector machine constructs a hyperplane or set of hyperplanes in a high- or infinite-dimensional space, which can be used for classification, regression, or other tasks like outlier detection. Intuitively, a good separation is achieved by the hyperplane that has the largest distance to the nearest training-data point of any class as shown in Fig. 3. It is widely used in various branches of pattern recognition. It was first proposed by Vapnik et al. [14]. It is especially suitable for high-dimensional and small-sample problems and has good promotion ability. The radial basis function (RBF) kernel is used in this work,

$$K(x, y) = \exp(-\gamma\|x - y\|), \gamma > 0 \qquad (7)$$

where γ is a kernel parameter.

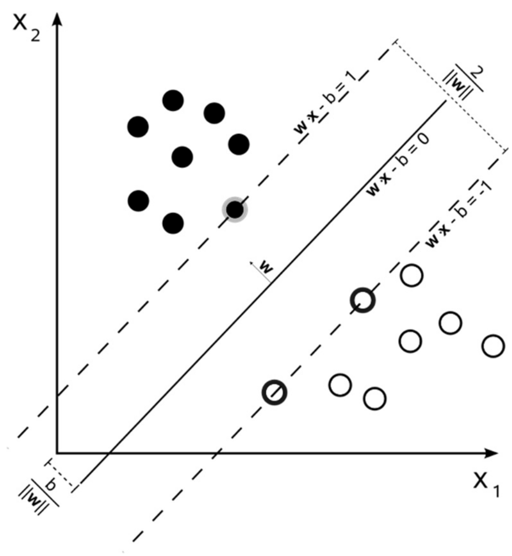

Figure 3. Maximum-margin hyperplane and margins for an SVM trained with samples from two classes. Samples on the margin are called the support vectors.

4. Experimental Results

The experimental process is as follows. First, in an automatic calibration of human face feature points, a square human eye region is extracted based on the positions of the two eye corners, the width of which is the actual width of the human eye. Then the grayscale correction is performed. Finally the scaling is performed to a size of 32×32 pixels, thus obtaining Sample sets and then extract feature vectors with PCA or Gabor. Experimental data obtained from CEW dataset [17]. We select a total of 2384 closed-eye samples and 2462 opened-eye samples. We randomly used 1500 closed-eye samples and 1500 opened-eye samples for training, and the other 1846 samples were used for testing. Fig. 4 shows some normalized left-eye samples, size 32x32, the first line of closed-eye samples, and the second line of opened-eye samples.

In our experiment, we compared the 6 combined models using 2 feature extraction algorithms and 3 classifier algorithms. The results of different models are shown in Table 1. From the Table 1, we found that Gabor features outperform PCA features in feature extraction of eye states. In the three classifiers, the Gabor features showed better prediction accuracy than PCA. For example, the classification effect of Gabor+NBC is 30.2% higher than that of PCA+NBC. In addition, by comparison, we also get better performance of SVM by comparing with KNN and NB. For example, the Gabor+SVM classification results are larger than Gabor+KNN and Gabor+NB with 3.4% and 1.9%, respectively. Therefore, we conclude that the Gabor+SVM algorithm achieves the most accurate classification result, which has a classification prediction with 93.1%.

Figure 4. Samples of normalized eye photos.

Table 1. Eye state classification results of different machine learning models.

Feature Types	*Classifiers*	*Accuracy*
PCA	KNN	0.833
PCA	NBC	0.610
PCA	SVM	0.795
Gabor	KNN	0.897
Gabor	NBC	0.912
Gabor	SVM	0.931

5. Conclusions

In this paper, we have especially aimed to support the applications of students' visual attention and participation for eLearning. This paper presents an eye state classification model based on machine learning algorithms. Experimental results show that this method achieves good performance. We systematically compare a variety of pattern classification models combining PCA, Gabor feature extraction algorithms and KNN, NB, SVM classification algorithms. By comparing these models, we found that the combination of Gabor and SVM has the best classification effect. The classification accuracy reached 93.1%, which has a significant value in real applications. In addition, we proposed a framework consisting of the automatic face detection and feature points detection based on current machine learning models as well. These models are labour-saving and provide more precise detection than manfully labelling [16]. Thus, we conclude that our proposed framework can not only obtain a satisfying performance in eye state classification, but also has practical value (labour-saving).

It is expected that the developments made during this study will motivate future research efforts. Cross-sectional research designs are also needed to extend the validity of the findings. The limitation of this study is that we only discuss the two states(opened and closed) of eyes on the students' visual attention. Although the study [15] shows that opened and closed eyes are the key factor of evaluating students' visual attention, the attention states of students is exactly influenced by many factors, such as frequency of wink. In the future, we can use the proposed framework to analyse another eye state like frequency of wink. Based on the proposed framework, a widely used assessment system for eLearning would be designed.

Acknowledgement
This paper is supported by the project of Jiangxi Provincial Educational Scientific Planning(16ZD016).

References
[1] Jeong H Y, Yeo S S. The quality model for e-learning system with multimedia contents: a pairwise comparison approach[J]. Multimedia Tools and Applications, 2014, 73(2): 887-900.
[2] Stanisavljevic Z, Nikolic B, Tartalja I, et al. A classification of eLearning tools based on the applied multimedia[J]. Multimedia Tools and Applications, 2015, 74(11): 3843-3880.
[3] Lim H, Lee S G, Nam K. Validating E-learning factors affecting training effectiveness[J]. International Journal of Information Management, 2007, 27(1): 22-35.
[4] Gómez-Aguilar D A, Hernández-García Á, García-Peñalvo F J, et al. Tap into visual analysis of customization of grouping of activities in eLearning[J]. Computers in Human Behavior, 2015, 47: 60-67.
[5] Leow F T, Neo M. Interactive Multimedia Learning: Innovating Classroom Education in a Malaysian University[J]. Turkish Online Journal of Educational Technology-TOJET, 2014, 13(2): 99-110.
[6] Cheon J, Lee S, Crooks S M, et al. An investigation of mobile learning readiness in higher education based on the theory of planned behavior[J]. Computers & Education, 2012, 59(3): 1054-1064.
[7] Jabbar S, Naseer K, Gohar M, et al. Trust model at service layer of cloud computing for educational institutes[J]. The Journal of Supercomputing, 2016, 72(1): 58-83.
[8] Malik K R, Ahmad T, Farhan M, et al. Big-data: transformation from heterogeneous data to semantically-enriched simplified data[J]. Multimedia Tools and Applications, 2016, 75(20): 12727-12747.
[9] Pearson K. LIII. On lines and planes of closest fit to systems of points in space[J]. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 1901, 2(11): 559-572.
[10] Bro R, Smilde A K. Principal component analysis[J]. Analytical Methods, 2014, 6(9): 2812-2831.
[11] Jian-Jiun Ding, Time frequency analysis and wavelet transform class notes, the Department of Electrical Engineering, National Taiwan University (NTU), Taipei, Taiwan, 2011.
[12] Altman N S. An introduction to kernel and nearest-neighbor nonparametric regression[J]. The American Statistician, 1992, 46(3): 175-185.
[13] Rennie J D, Shih L, Teevan J, et al. Tackling the poor assumptions of naive bayes text classifiers[C]//Proceedings of the 20th international conference on machine learning (ICML-03). 2003: 616-623.
[14] Cortes C, Vapnik V. Support-vector networks[J]. Machine learning, 1995, 20(3): 273-297.
[15] Farhan M, Aslam M, Jabbar S, et al. Multimedia based qualitative assessment methodology in eLearning: student teacher engagement analysis[J]. Multimedia Tools and Applications, 2018, 77(4): 4909-4923.
[16] Sagonas C, Tzimiropoulos G, Zafeiriou S, et al. 300 faces in-the-wild challenge: The first facial landmark localization challenge[C]//Computer Vision Workshops (ICCVW), 2013 IEEE International Conference on. IEEE, 2013: 397-403.
[17] Song F, Tan X, Liu X, et al. Eyes closeness detection from still images with multi-scale histograms of principal oriented gradients[J]. Pattern Recognition, 2014, 47(9): 2825-2838.