

Millennial Filipino Student Engagement Analyzer Using Facial Feature Classification

R Manseras¹, F Eugenio², and T Palaoag³

¹University of the Immaculate Conception, Davao City

Isabela Colleges Incorporated, Cauayan City, Isabela

University of the Cordilleras, Baguio City

¹roymanseras@gmail.com ²franciscojreugenio102683@gmail.com

³tpalaoag@gmail.com

Abstract. Millennials has been a word of mouth of everybody and a target market of various companies nowadays. In the Philippines, they comprise one third of the total population and most of them are still in school. Having a good education system is important for this generation to prepare them for better careers. And a good education system means having quality instruction as one of the input component indicators. In a classroom environment, teachers use facial features to measure the affect state of the class. Emerging technologies like Affective Computing is one of today's trends to improve quality instruction delivery. This, together with computer vision, can be used in analyzing affect states of the students and improve quality instruction delivery. This paper proposed a system of classifying student engagement using facial features. Identifying affect state, specifically Millennial Filipino student engagement, is one of the main priorities of every educator and this directed the authors to develop a tool to assess engagement percentage. Multiple face detection framework using Face API was employed to detect as many student faces as possible to gauge current engagement percentage of the whole class. The binary classifier model using Support Vector Machine (SVM) was primarily set in the conceptual framework of this study. To achieve the most accuracy performance of this model, a comparison of SVM to two of the most widely used binary classifiers were tested. Results show that SVM bested RandomForest and Naive Bayesian algorithms in most of the experiments from the different test datasets.

1. Introduction

Millennials, also known as Generation Y, has always been a buzzword to every news articles and target market of most of the product endorsements. The term Millennials refers to a person reaching adulthood in the early 21st Century [1]. They are known to be electronic-oriented and internet-dependent people.

In the Philippines, Filipino millennials comprise one third of its population. A large number of these tech-savvy individuals are currently still in school and soon will they be movers of the economy in the future.

Education is the basic foundation of any individuals who want to survive as adults in this world. The quality of instruction delivery is an important element in achieving a better education system and this includes student engagement as its binding factor [2].

Student engagement determines a student's level of attention, curiosity and interest in the class. And for every teacher, improved engagement is part of his or her top priorities [3]. Student engagement has three (3) incorporating components: behavioral, cognitive, and emotional. A student's conduct and



participation to co-curricular and extra-curricular activities is under behavioral engagement. A student's ability to understand key concepts and skills needed for learning is what is being measured under cognitive engagement. And how students react towards the people that surround them is under emotional engagement. This paper focused on student engagement and uses facial features to find out what is the current affect state of the students.

The incorporation of technology in education is among the new approaches that are being applied to prepare students for life-long learning. In the same way, the field of computing is leaning towards the implementation of computer vision-based applications. Affective computing combined with computer vision plus education, with a focus on student engagement, is a combination that is suitable for dealing with student affect states. This, along with the use of some tools (sensors, microphones, camera, etc.) [4], can contribute to improving how teachers analyze students' engagement states through the evaluation of facial features.

Machine-learning algorithms have already been used and effectively applied, with improved percentage of accuracy, in the real-time detection of facial expressions [5]. Studies have also been made that focused on detecting learning-centred states such as confusion, excitement, flow, frustrations [6], informed, inspired, persuaded, sentimental and amused [7]. There are researches that used Facial Action Coding System (FACS) in successful implementations of emotion detection through facial features. In a study by Ekman et. al, they conducted a thorough examination of the relationship of emotions through facial features [5] called Action Units (AUs). These facial features (forehead, nose, mouth, eyes, etc.) [5] were linked to certain emotions and a connection was found on some of these AUs with student engagement.

In this paper, the researchers aim to integrate affective computing and computer vision to aid teachers in gauging quality of instruction delivery by analyzing Filipino millennials' affect state, specifically student engagement during class discussion, through facial feature classification.

Teachers could best identify student engagement during class discussion. Hence, the researchers would like to prove that initial dataset labeled by teachers would be effective as basis in classification of student engagement by the proposed Predictive Classification Model.

2. Methodology

Figure 1 illustrates the methodology used by the researchers. The model served as a guide for the researchers to perform the classification and identification of the student engagement.

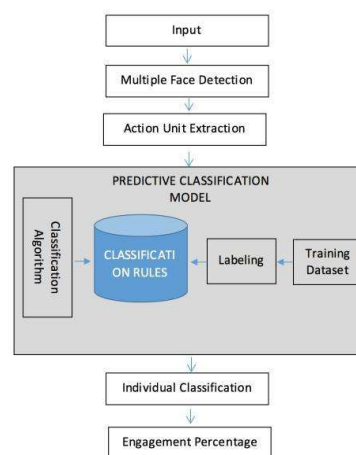


Figure 1. Conceptual Framework

2.1. Input

The input (see figure 1) would be a still image captured through a camera in a classroom environment where students are front facing the teacher during a class discussion. To address biased image capture of the class' affect state, system setting could be set to capture an image randomly between the specified time range within class discussion. Captured image will then be loaded for the multiple face detection.

2.2. Multiple Face Detection

To optimize the detection of multiple faces of students in class, an advanced machine-learning framework has been used for this experiment. The researchers used Face API which is one of the products of Microsoft Cognitive Services. The latter utilizes Microsoft Cognitive Toolkit (CNTK) as its backend. It is an open-source toolkit for deep learning algorithms. This technology employs computational network (CN) which is a framework consisting of multiple components like deep neural networks (DNNs), convolutional neural networks (CNNs) and others. It has been known that this toolkit evaluates deep learning algorithms faster than other toolkits.

The Face API of Microsoft Cognitive Services offers several features like face detection, face verification, face identification, similar face searching, and face grouping. Its face detection capability could detect multiple faces in a single still image.

2.3. Feature Extraction

After all possible faces, have been detected, the system will then apply the individual faces for identifying the various facial Action Units. These AUs will then be used as parameters for the classifier.

The system will utilize OpenFace, an open source program interface that is capable of AU recognition, facial landmarks detection, head pose estimation and gaze estimation in low hardware settings. Refer to Table 1 for the different Action Units described.

Table 1. Action units detected by OpenFace

AU	Description	AU	Description
AU1	Inner brow raiser	AU14	Dimpler
AU2	Outer brow raiser	AU15	Lip corner depressor
AU4	Brow lowerer	AU17	Chin raiser
AU5	Upper lid raiser	AU20	Lip stretched
AU6	Cheek raiser	AU23	Lip tighter
AU7	Lid tighter	AU25	Lips part
AU9	Nose Wrinkler	AU26	Jaw drop
AU10	Upper lip raiser	AU28	Lip suck
AU12	Lip corner puller	AU45	Blink

2.4. Classifier Model

To create this binary classification model, authors utilized the extracted Action Units label (true or false) and used them as parameters for classification using Support Vector Machine. In this experiment, 2 machine learning algorithms were tested and compared to prove that SVM is the most appropriate for this proposed prototype.

- Support Vector Machine (SVM)*: SVM has been proven to be effective in some known facial recognition systems like OpenCV and CERT [9]. It is an algorithm that classifies new dataset based on the given supervised learning. SVM yields the best hyperplane that has the largest minimum distance to the given data.
- Naive Bayes*: Naive Bayesian algorithm is one of the widely-used classifier by various researchers because it is easy to build and applicable to large dataset. This algorithm is based on Bayes' theorem with independence assumptions between predictors.
- Random Forest*: It is a framework consist of various methods for classification and regression and other types. This framework works by establishing a number of decision trees at training time and display the class or mean production for the individual trees. Some of its best features include high accuracy even on large dataset and it could handle multiple variables without requiring reduction on attributes in high dimensional data.

The performance of these three algorithms were estimated with cross-validation. To measure the accuracy of the trained classifier, 10-fold cross-validation (CV) was used. This process will divide the given training dataset into smaller subsets. Table 2 shows that Support Vector Machine gained well in

classifying data instance as compared to the Naive Bayes and RandomForest. These results only indicate that SVM best performs on the validation data and does not imply accuracy on unseen dataset.

Table 2. 10-fold cross validation of the training dataset

Algorithm	Precision	Recall	F-Measure
Naive Bayes	0.823	0.843	0.833
SVM	0.842	0.880	0.861
RandomForest	0.871	0.839	0.855

2.5. Data Selection

Training and Validation dataset as input to the classifier model will use the Extended Cohn-Kanade (CK+) AU-Coded facial expression database. This set of images consist of 123 subjects with 593 sequences [10].

In this prototype, the model will initially use labelled training dataset with the aid of human experts, in this case, two teachers as labellers. Data label for each instance could either be Engaged or Not Engaged.

Indicators used for label 'Engaged' student as discovered in [8], include expressions that denote concentration, gestures and excitement or presence of any of the following AUs: AU7(Lid Tighter) +AU12(Lip Corner Puller), AU5(Upper Lid Raiser), AU25(Lips Part) and AU26 (Jaw Drop); otherwise, such instance will be labelled 'Not Engaged'.

Certain procedure was applied in selecting the training dataset for the binary classification to yeild better predictive model. If the assigned label differs from the 2 labelers (e.g. labeler1 assigned Engaged and labeler2 assigned Not Engaged), the image will be discarded. This process reduced the original dataset from 593 to 499 instances.

Final test data include 3 different datasets. The first 2 datasets were captured images from 2 classes with 47 and 13 visible faces respectively. Last test dataset consists of combined instances from the first 2 datasets.

2.6. System Feedback

After going through the whole process of image capture to classification, the final system feedback to the user would be a percentage of labelled Engaged students.

2.7. Accuracy Metrics

To measure the quality of the binary classifier with respect to the test dataset, Precision, Recall and F-Measures were acquired based on true positives(TP), false positives(FP) and false negatives(FN). F-measure is the weighted average of Precision and Recall, where score reaches its best value at 1 and worst at 0. Refer to Equation 1.

$$\text{F-Measure} = 2(\text{recall} \cdot \text{precision}) / (\text{recall} + \text{precision}) \quad (1)$$

3. Results and Discussions

3.1. Face Detection and Feature Extraction

Multiple face detection using Face API yielded a considerable result. Out from the 47 subjects of the first image input, only 32 were detected (see figure 2) or 68%. And from 13 students present in the image for the second input, Face API detected 10 faces or equivalent of 77%. With an overall average of 73%. The researchers observed that the following factors are causes for non-detection of faces: (a) Non-front facing position; (b) Incomplete facial features required for face detection (e.g. eyes blocked by other objects); (c) Too far or too close from the camera causing very blurry or faded face image.



Figure 2. Screen capture from actual face detection using Face API

Moreover, facial Action Units were successfully extracted by OpenFace even with the very small image output cropped from the detected images of Face API with 64x64 image size (see figure 3).

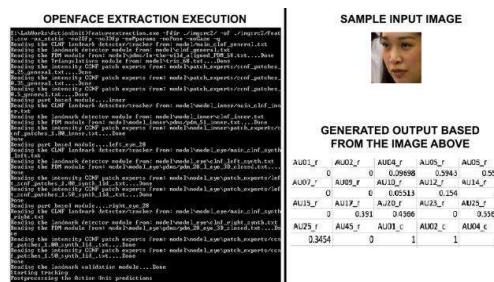


Figure 3. OpenFace Execution

3.2. Algorithm Performance with Test Datasets

The following explains the experiment results of the different algorithms, with emphasis on F-Measure, used with the final test datasets shown in tables 3,4,5.

Table 3. Accuracy of Algorithms in dataset1 with 32 instances

Algorithm	Precision	Recall	F-Measure
Naive Bayes	0.8	0.421	0.552
SVM	0.833	0.526	0.645
RandomForest	0.625	0.526	0.571

Table 4. Accuracy of Algorithms in dataset2 with 10 instances

Algorithm	Precision	Recall	F-Measure
Naive Bayes	1	0.6	0.75
SVM	0.75	0.6	0.667
RandomForest	0.4	0.4	0.4

Table 5. Accuracy of Algorithms in dataset3 with 42 instances

Algorithm	Precision	Recall	F-Measure
Naive Bayes	0.846	0.458	0.595
SVM	0.813	0.542	0.65
RandomForest	0.571	0.5	0.533

RandomForest algorithm lags behind the 2 other classifiers on all datasets with an average of 0.5 F-Measure value. Moreover, NaiveBayesian algorithm performed as the best classifier based on dataset 2 but remained on the second spot after increase in number of instances. Furthermore, SVM ranked second on the dataset2 with fewer instances but as the number of instances increased, so as its accuracy performance.

4. Conclusion and Future Works

The authors hypothesized that the use of the human experience by the teachers would best create a model based on certain indicators to detect student engagement. This assumption was proven through the F-measure achieved by the binary classifier model using the labelled training dataset. SVM achieved a very high accuracy rate based on F-measure on most of the test dataset experiments. These results imply the robustness of this algorithm as it achieved a very high precision rate in detecting correct classification in most of the instances.

Furthermore, other components of the framework need further studies to best improve this proposed prototype. In the case of multiple face detection, it is best to compare the Face API performance to other available face detection system specifically those frameworks that could address the current limitations of the CNTK. Moreover, with regards to the binary classifier model, using SVM has proven to be effective on classifying student engagement. However, authors strongly recommend that additional faces for training dataset must be sought to improve better classification and the testing of the prototype must be set to various classroom environments to assess the efficiency of the study in different test datasets.

References

- [1] — 2017 *What is Millennials (Millennial generation)? - Definition from WhatIs.com* [Online] Available: <http://whatIs.techtarget.com/definition/millennials-millennial-generation>.
- [2] Kecojovic Z 2012 Perspectives on Educational Quality: Illustrative Out- comes on Primary and Secondary Schooling in the Netherlands *CEPS J.: Center for Educational Policy Studies J.* **2(2)** p 145.
- [3] Abbott S 2017 [Online] Available: <http://edglossary.org/student-engagement/>.
- [4] Banafa A 2016 *What is Affective Computing?* [Online] Available: <https://www.bbvaopenmind.com/en/what-is-affective-computing/>.
- [5] Al-Alwani A 2016 Estimation of Mental States Using Facial Features to Improve A Students Performance In E-Learning *Academics World52nd Int. Conf.*
- [6] Grafsgaard J, Wiggins J, Boyer K, Wiebe E and Lester J 2016 Automatically Recognizing Facial Indicators of Frustration: A Learning-Centric Analysis *Humaine Association Conf. on Affective Computing and Intelligent Interaction*.
- [7] McDuff D 2016 Discovering Facial Expressions for States of Amused, Per- Suaded, Informed, Sentimental and Inspired *18th ACM International Conference on Multimodal Interaction* (ACM) pp 71-5.
- [8] Nyman R 2015 Indicators of Student Engagement: What Teachers Notice During Introductory Algebra Lessons *Int. J. for Math. Teaching and Learning*.
- [9] Littlewort G, Whitehill J, Wu T, Fasel I, Frank M, Movellan J and Barlett M 2011 The Computer Expression Recognition Toolbox (CERT) *Automatic Face and Gesture Recognition and Workshops (FG 2011)* (IEEE) pp 298-305.
- [10] Lucey P, Cohn J, Kanade T, Saragih J, Ambadar Z and Matthews I 2010 The Extended Cohn-Kanade Dataset (Ck+): A Complete Dataset for Action Unit and Emotion-Specified Expression *Computer Vision and Pattern Recognition Workshops (CVPRW)* (San Francisco, USA: IEEE) pp 94-101.