Environment. Technology. Resources. Rezekne, Latvia
Proceedings of the 14th International Scientific and Practical Conference. Volume 2, 84-88

Automation of Feedback Analysis in Asynchronous E-learning

Rohit
Faculty of Engineering
Rezekne Academy of Technologies
Rezekne, Latvia
rohitpruthi55@gmail.com

Peter Grabusts
Faculty of Engineering
Rezekne Academy of Technologies
Rezekne, Latvia
peteris.grabusts@rta.lv

Artis Teilans
Faculty of Engineering
Rezekne Academy of Technologies
Rezekne, Latvia
artis.teilans@rta.lv

Atis Kapenieks
Distance Education Study Centre
Riga Technical University
atis.kapenieks@rtu.lv

***Abstract*. With the recent hit of the Pandemic study process is shifted to E-learning. Measuring the actual progress of a student in Asynchronous e-learning because of machine-human interaction and feedback is also considered a primary issue in this area. With the rapid development of artificial intelligence, computers can capture surroundings. Image processing is a rising technique in Artificial Intelligence (AI). Recognition of an individual's emotion helps identify the person's inner state. It is easy to measure a student's feedback about the session by doing it. The main functionality of this project is to capture the student's enough frames during the study session and provide the analysis of the average emotions to the administration panel. This prototype was used on 10 minutes of lecture to capture the emotion. The primary goal of this proposed system is to capture the learner's emotion at a specific interval during the e-learning session and provide the feedback of it to the instructor.**

Keywords: Computer vision, Convolutional neural network, E-learning, Emotion Feedback, Facial Expression.

I. Introduction

The word “feedback” is very familiar. Feedback can help to improve the quality of a session in many ways. Most of the study process continues with the traditional form of feedback. In the study process, the trainee immediately always gives the student feedback. It is usually given in facial expressions, gestures, or comments. The remaining problem in a study session is the student's genuine feedback after the completion of the session. It is easy to ask the learners to provide feedback in a traditional classroom study session. With the recent pandemic hit, the study process is shifted to digital sources known as E-learning [1]. Feedback is considered a problematic issue in higher education [2]. There is no proper definition of E-learning. Some E-learning reports went on to ask: Is E-learning creating a virtual environment to provide campus-based education? [3].

E-learning refers to communication technologies and information to enable the learning and teaching process. An E-learning system has a few of its characteristics. To begin with, it proposes a multimedia environment. In addition, it supports collaborative communication, whereby users have total control over their learning situation. E-learning is mainly divided into two parts – Synchronous and Asynchronous. Further, it is divided into more parts. Synchronous E-learning Is the interaction between human to humans through digital resources. It is also known as real-time learning. It creates a virtual classroom with the presence of the instructor or teacher. In this simultaneously, communication is possible between teacher-student and student-student. The most common mode used for this type of e-learning is video conference or audio conference. Asynchronous E-learning is the exact opposite of Synchronous E-learning. In this E-learning, leaner can adjust the learning time according to their schedule. There is no establishment of a virtual classroom between teacher and student in this E-learning. This E-learning can also be described as Self-study. The media used in this learning process is multimedia disks, external drives, or downloading forums. This learning is more popular than

Print ISSN 1691-5402
Online ISSN 2256-070X
https://doi.org/10.17770/etr2023vol2.7285

synchronous learning as learners can study according to their schedules [4]. Talking about e-learning session feedback in synchronous interaction takes place between human-human through digital resources, and taking feedback is not such a complicated task. As feedback Is considered the main factor in determining the success rate of a session, i.e., Was the trainee able to provide accurate information according to the session? Was the learner able to acquire all the information related? Is the learner ready for the next session? Identifying real-time feedback or feedback in synchronous e-learning depends on a few factors, i.e., body language, facial expression, or sentiments.

For human-to-human interaction, facial expressions are essential since expressions describe the current inner state and the actual reaction to the ongoing session [5]. Asynchronous E-learning is the interaction between a machine and a learner. It is not possible to get feedback in a live session. So, it is the possibility for Constructive feedback after the session. So genuine or accurate feedback from learners in asynchronous e-learning is still a problematic issue. Measure a student's success in the current session is almost impossible for the instructor or teacher. Facial expression also plays a vital role in the feedback area of any activity. This paper came up with the idea to capture human facial expressions via machine to find out the inner state of the learner during the session. Nowadays, with the help of AI, machines are smart enough to make decisions. AI is a sub-discipline of computer science. It is the most popular topic in today's time. AI takes over almost every field of humans. As their human exposure increased, the interaction became smoother and more natural. Nowadays, machines can capture environment surroundings through cameras and sensors. During the interaction between machines and humans, machines can capture human behaviour. In the past few years, deep learning algorithms have proven very successful. To better serve human affective computing, emotion detection is vital for machines. Aim of the study is to train a deep convolutional neural network with images of static facial emotions and use it as part of the software to detect their inner state during an asynchronous e-learning session. These results can improve the communication between students and lecturers and help improve how a teacher or course designer can make it more accessible or more attractive for a student [6].

II. Human Facial Emotion

The role of human facial expression in feedback plays a very vital role. Facial expression helps to understand the inner state of an individual on a deeper level [7]. To measure emotional responses, most research focused on recognizing the basic emotions. There are six different facial emotions. It is the most common feature of humans to react to any task. The primary six expressions are, i.e.- “Happy”, “Sad”, “Angry”, “Surprise”, “Neutral”, “Disgust”.

Above written expressions are the most common for every face. It is also customary to understand any human by its manifestations. In different types of emotions, human face muscles make various movements. A simple difference is visible in the Fig. 1.

Fig. 1. Different states of human facial expression (source: [11]).

The image of the left side presents a different expression from the image of the right side. It shows how a person reacts to different situations. Normal taking this concept to a study room session. In the case of Synchronous E-learning, Its interaction between human-to-human faces. So recognizing a student's expression in session can be easily identified by the teacher. If a student feels happy, it is understandable that they are learning something new and are pleased with the session. Every expression has its characteristics. i.e.

Sad – It indicates a student is not mentally ready for this session.

Happy – It indicates a student is enjoying the session.

III. Neural Networks

It is an approach inspired by the human brain and teaches a computer to process data. It creates an adaptive system in that computer improves themselves by learning from their mistakes. It is a significant branch of AI (see Fig. 2).

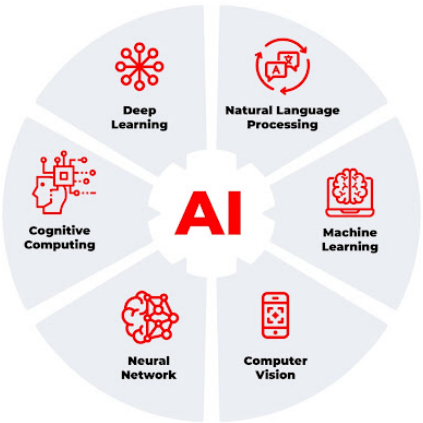

Fig. 2. Branches of AI (source: https://10pie.com).

The main characteristic of a Neural Network (NN) are following. It helps to solve problems and make intelligent decisions without human assistance. NN analyze raw data profoundly and reveal new insights for which they might not have been trained. In today's time, usage of the NN is in many industries, i.e., medical diagnosis, target marketing, financial predictions etc. [8]. NN have different types, but this paper focuses explicitly on Convolutional

Neural Networks (CNN). CNN is beneficial for image classification because it can extract features from images for recognition and classification. The main advantage of CNN is that it automatically detects the essential features without human supervision. CNN aim to use information between the pixel of an image [9]. CNN is primarily used to classify images. CNN can recognize street signs, animals, and human faces using many algorithms. The correlation between CNN and feedback is Human behaviour. As mentioned above, human facial expressions are a very critical characteristic. With the use of CNN, it is effortless to capture human facial expressions from a live feed during the session. Human facial has a limited number of expressions. i.e., Happy, Sad, Angry, Neutral, disgusted, Surprise. CNN can identify these expressions. Further, these results can be used to improve the quality of education. The parameters used to train the CNN are mentioned in the Implementation chapter.

Fundamentally, CNN architecture combines two main parts (see Fig. 3).

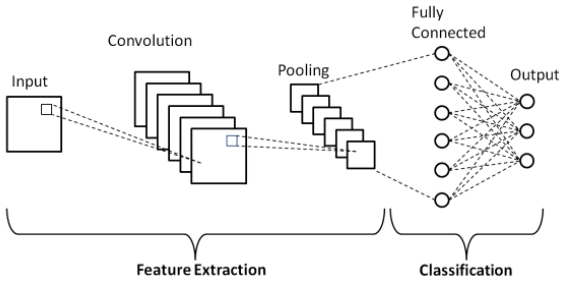

Fig. 3. Architecture of CNN (source: https://www.upgrad.com/blog/basic-cnn-architecture/).

Feature extraction – this layer slides over the input data and identifies the various information. This process is called feature extraction.

Classification - this layer utilizes the output from the process and predicts the image class based on the feature extraction process.

Simple example of objection detection by CNN see in Fig. 4.

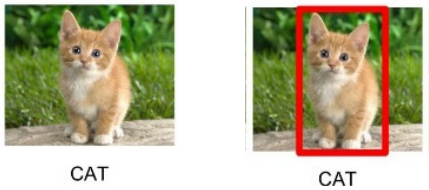

Fig. 4. Objection detection by CNN (source: https://www.alibabacloud.com/blog/).

The first image is the sample image used to check whether that cat is available or not in the image. This image was feed in to the trained CNN to check the cat is present in this picture. After putting visualization, it can detect a cat.

IV. MATERIALS AND METHODS

The complete process of automation of feedback analysis is divided into three modules.

A. Frame extraction with face detection

The first step is to extract the frames from the live feed or video with face detection. A live feed is a collection of frames per second, so the extraction is done by writing a simple frame extraction program in python language. Many approaches, methods, and techniques for detecting a human face in a digital image. In the 1960s, Woodrow Wilson Bledsoe manually implemented recognition of a face. Detection of the object in the digital image is connected to computer vision.

The main aim of the detection of face algorithm is to determine whether there is a face or not in the given image. Face detection can be done via different methods, i.e. knowledge-based, appearance-based, feature-based, and template-matching using a Deep Neural Network (DNN). However, detection of the face could be done by DNN, but a predefined computer vision library named "OpenCV" was used to detect the face. It is an open-source library. This library contains around 500 functions that span many areas of computer vision, and detecting an object is one of them [10]. This library helps to use computer vision algorithms straightforwardly. The primary advantage of using OpenCV is highly optimized and available on all popular platforms. Haar cascade classifier algorithm is used to detect the face in the given video.

The main reason behind using computer vision instead of using a NN for detection is that it can detect faces during the live feed. So the frame will extract only based on the face detected in that frame. If the program does not detect a face in a frame, it will take another frame instead of without a face and store it in the user's unique directory. A course designer can define how many total frames are required to measure the learner's progress. In this paper, the author assumes the session is about 10 minutes and extracts frames after every 10 seconds. The total number of frames will be 60.

B. Analysis of facial emotion

The extracted frame result from the Haar cascade classifier algorithm will be directly forward to the trained CNN to analyze the emotion. The program will run a loop to check emotion on every single frame. Results from every single frame will be stored in relevant information holding variable to the emotion.

C. Analysis of result and improvement process

The teacher will analyze the results at the end of the session; after the analysis, the teacher will take the necessary steps (see Fig. 5). In case the teacher feels the student needs to improve the previous session, then the student will retake the same session with the improvement of content; on the other hand, if the teacher feels improvement in the content of the next session can take the result for the overall course to a satisfactory level. In that case, the content of the next session will be updated before the session. The improvement process will be repeated until the result of the current session reaches a satisfactory level.

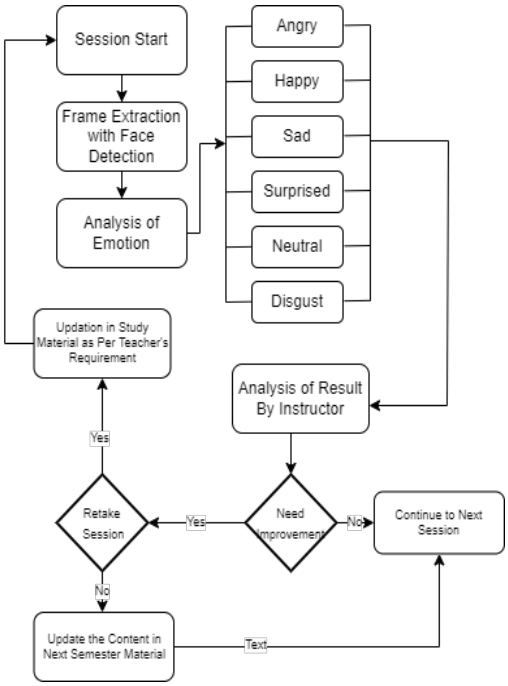

Fig. 5. Structure of the emotion analysis system.

V. IMPLEMENTATION

A. Dataset

Dataset is a set of collections used to train a network for future predictions. Dataset can be in graphical form as well as in tabular form. In tabular form, each row contains pixel values of the image or corresponds to a given data record, and each column represents a particular variable. It can be collections of files or documents.

The dataset used in this research contains 30000 images with the size of 48*48 mm and a bit-depth of 8 bits. The name of the dataset is FER2013, and this dataset is available on the internet [11]. The dataset credit goes to Pierre-Luc Carrier and Aaron Courville. The images used are of people representing six different emotions. For the training of the CNN, 90% dataset was used, and the remaining dataset was used for testing purposes. The sample of used images see in Fig. 6.

These images represent emotional happiness. Similarly, all emotions have such kinds of images. For each emotion, around 5000+ images were used for training purposes, which helped maximize the model's accuracy.

B. Training Process

The training process of CNN basically making a layers structure to feed the data for learning process. Training process can also be called learning process. In this process model use to learn from the data samples. The layer structures which is used to train model is given in Fig. 7.

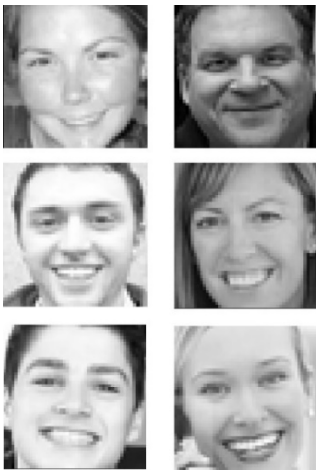

Fig. 6. Sample dataset images (source: [11]).

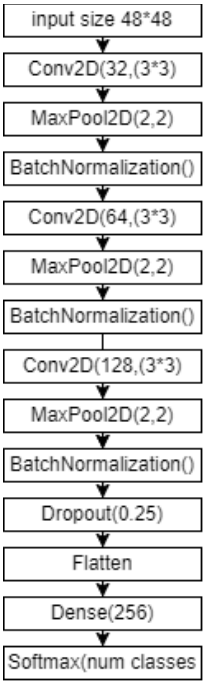

Fig. 7. Structure of CNN.

The current accuracy achieved with the CNN structure with the above mentioned dataset is 72%.

C. Frame extraction from live feed

Live feed is a massive volume object with a piece of insensitive information and high redundancy. It has a complex structure with the scene, shot or frames. One of the fundamental units in structure analysis is the Key-frame extraction. The first step is to extract the frame from the live feed. Extracting frames is a separate process in this project. It doesn't use any algorithm used for feedback analysis. Every second is very important to analyze the learner's emotions during the event. Thus, a separate set of commands is defined for frame extraction, which allows extracting the frames as per the course designer's request. This program was tested with a sample session of 10 minutes, and some of the extracted frames from live feeds are following (see Fig. 8).

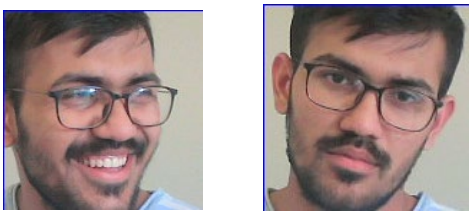

Fig. 8. Extracted frames from live feed (source: personal archive).

Analysis of emotion

The programs will run in a loop to detect the emotion. Every emotion output will be stored in different variables to calculate the learner's average emotion status during the session. After detecting emotion from all frames, the result will be present in the graphical bar on the teacher's panel to measure student progress.

Result

After calculating the learners' emotional states, in the below bar, the results are presented (see Fig. 9).

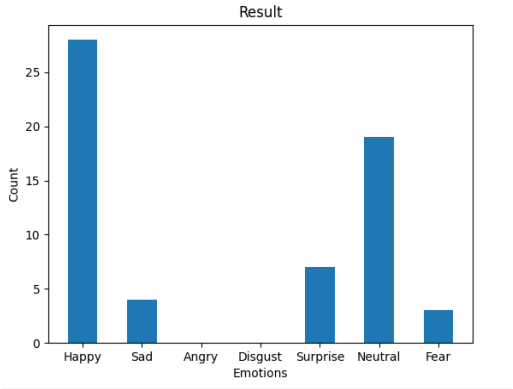

Fig. 9. Bar graph showing emotion states of learner.

These results are from 10 minutes of demo sessions with 60 frames. These results will be visible only to the teacher. So for them, it's straightforward to measure the learner's success during the session if we talk about the above result. It is effortless for a teacher to understand how the session went without the presence. As a result, shows max states of emotion and was happy in the session. In simple words, the learner enjoyed the session.

VI. Learning Analytics

Learning Analytics is the new trend in digital learning. The popular explanation of Learning Analytics includes the "measurement, collection, analysis and reporting of data about learners and their contexts". The Learning Analytics aims to optimize the learning. Most of the Learning Analytics solutions are using the data sets similar to traditional education. The face emotion detection has the great potential to extend and enrich Learning Analytics data set and usability.

The next step we see the complimentary extension of Learning Analytics data set with Knowledge Acquisition Monitoring [12]. The both technologies could create the fast feedback for update of learning content and delivery approach.

VII. Conclusions

Although automatic detection of the face in today's time is not so tricky, detecting a tilted face is a complex task. The most challenging part is to train the CNN repeatedly to get better results and accuracy to detect the emotion more accurately. These results will be used as feedback from the student in live sessions. It will help make the content much more exciting and accessible for students.

In future, we will develop this project by adding keynotes to a study session to measure the progress more accurately. We will retrain the model through with transfer learning to achieve more accuracy. This project will be implemented into a Moodle system also. Apart from this, we will develop this project with more actions of the user, such as sitting pose or voice input, to measure the student's learning progress in a much more accurate way.

References

[1] H. Alodan, "E-Learning transformation during the covid-19 pandemic among faculty members at princess nourah bint abdul rahman university," Utopía y Praxis Latinoamericana, vol. 26, no. 2, pp. 286-302, 2021.

[2] M. Mamoon-Al-Bashir, M. Rezaul Kabir and I. Rahman, "The Value and Effectiveness of Feedback in Improving Students Learning and Professionalizing Teaching in Higher Education," Journal of Education and Practice, vol. 7, no. 16, pp. 38-40, 2016.

[3] R. Maltese, "How To Create a Virtual Classroom," 2021. [Online]. Available: https://corp.kaltura.com/blog/creating-a-virtual-classroom. [Accessed March 1 2023].

[4] A. Perveen, "Synchronous and Asynchronous E-Language Learning: A Case Study of Virtual University of Pakistan," Open Praxis, vol. 8, no. 1, pp. 21-39, 2016. DOI: http://doi.org/10.5944/openpraxis.8.1.212

[5] S. Rukavina, S. Gruss, H. Hoffmann and H.C. Traue,"Facial Expression Reactions to Feedback in a Human-Computer Interaction - Does Gender Matter?". Psychology, Vol.7 No.3, 2016. DOI: http://dx.doi.org/10.4236/psych.2016.73038

[6] M. Binu, "The role of feedback in classroom instruction," The Journal of ELTIF, 2020.

[7] "Facial expressions," The Emotional Intelligence Academy. Online]. Available: https://www.eiagroup.com/knowledge/facial-expressions/. [Accessed March 1 2023].

[8] "What is a neural network," Amazon Web Services. [Online]. Available: https://aws.amazon.com/what-is/neural-network/. [Accessed March 1, 2023].

[9] D. Stutz, "Understanding Convolutional Neural Networks," 2014. [Online]. Available: https://davidstutz.de/wordpress/wp-content/uploads/2014/07/seminar.pdf. [Accessed March 1, 2023].

[10] G. Bradski and A. Kaehler, Learning OpenCV. O'Reilly Media, 2008.

[11] "FER2013 Dataset," [Online]. Available: https://datasets.activeloop.ai/docs/ml/datasets/fer2013-dataset/ [Accessed March 1, 2023].

A. Kapenieks, I. Daugule, K. Kapenieks, V. Zagorskis, J. Kapenieks Jr., Z. Timsans and I. Vitolina, "TELECI Approach for e-Learning User Behavior Data Visualization and Learning Support Algorithm, "Baltic J. Modern Computing, Vol. 8, No. 1, 129-142, 2020. https://doi.org/10.22364/bjmc.2020.8.1.06