Detection of Subject Attention in an Active Environment Through Facial Expressions Using Deep Learning Techniques and Computer Vision

Chapter · June 2020

DOI: 10.1007/978-3-030-51041-1_43

Naqash Gerard
National University of Sciences and Technology

Ahmed Husnain Johar
National University of Sciences and Technology

Umer Asgher
National University of Sciences and Technology

Detection of Subject Attention in an Active Environment Through Facial Expressions Using Deep Learning Techniques and Computer Vision

Naqash Gerard(✉), Talha Yousuf, Ahmed Husnain Johar, Umer Asgher, Imran Malik, Adnan Ul Hasan, and Faisal Shafait

National University of Sciences and Technology (NUST), Islamabad, Pakistan
{ngerard.msite18seecs,malik.imran,adnan.ulhassan,
faisal.shafait}@seecs.edu.pk,
{13bemetyousuf,ahmedhusnain.pg}@smme.edu.pk,
umer.asgher@smme.nust.edu.pk

Abstract. This research aims for investigation of workers in an industrial environment and can be used as an alternate for monitoring an attention of operator in real-time. Detection of attentiveness and non-attentiveness of people working in an industry could help to identify the weaknesses and strengths of any industrial organization. Human factor is the main and the most critical part of any industrial organization. As a special case, we have established how to detect student attention in the classroom using deep learning techniques along with computer vision. Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) are used to extract the percentage of attentiveness and non-attentiveness of students based on the student emotions in the classroom. We used the FER-2013 data set for this paper. As per the study, human has finite number of emotions. So, it is easy if we include some emotions in an attentive (Happy, Anger, Surprised and Neutral) domain and some emotions in non-attentive (Sad, Fear and Disgust) domain. This will help the teacher in a way that he can easily evaluate his class attentiveness. On another side, it is also the evaluation of the teacher's teaching methodology because if the students are engaged in his lecture it means his teaching methodology is good and if most of the students are not engaged then the teacher needs to revise his methodology of teaching in order to engage his class during the lecture.

Keywords: Human emotions detection · Emotion · LSTM · CNN · Computer Vision · Deep Neural Networks · Attention detection

1 Introduction

Technology has proven its importance in the field of Education. As the advancement in the technology, people always say that it's mandatory to integrate technology in the classrooms. Technology has a great impact during the learning of student. Students can use iPad, smart devices and laptops during their lectures instead of bringing a bag full of books and notebooks. Since the last decade, Machine Learning, Computer Vision

H. Ayaz and U. Asgher (Eds.): AHFE 2020, AISC 1201, pp. 326–332, 2021.
https://doi.org/10.1007/978-3-030-51041-1_43

and Artificial Intelligence has totally changed the perception of the technology geeks. Things that were once just an idea are now being developed. Like Siri by Apple, Google Assistant, Amazon Echo and many more devices and application which are built on the concepts of Machine Learning and Artificial Intelligence.

Student's attention plays an important key role towards learning. If student is attentive in the lecture, he will learn more of the content and understand the content. Students who are attentive make a psychological deal in learning and they try to understand and learn what they study during lecture [1, 2].

Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) network has been used to check the student attention in the classroom with the better accuracy. We have used state of the art dataset which is FER2013 [3] to detect the attentiveness of student through the facial emotions.

2 Literature Review

Attention and behavior of a student are the main and foremost part while attending a lecture. When the student is studying his behavior matters a lot because he must give proper attention to the specific subject.

2.1 Image Classification

Image classification is defined as, "A process that is used to classify an image based on its content or the visual appearance." The behavior of a student could be judged using the image classification techniques [13]. Image classification could be helpful in any domain of detection [10, 12]. If a student is attentive, he will put most of his effort to study and understand the subject more clearly. On the other side, if the student is not engaged most of the time, he gets himself busy in different things [4].

2.2 Neural Networks

A neural network is a set of algorithms that are capable of recognizing patterns and are inspired from human brain. The basic structure of a neural network is that it takes an input and applies certain mathematical operations on it and generates an output of a system. The layers on which the mathematical operations are applied is known as hidden layer. There are multiple type of neural networks but the approach we have used is Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM).

2.2.1 Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) is deep learning algorithm that is used to process visual imagery. The application scope of CNN covers medical image analysis, image and video recognition, and natural language processing [14, 15] etc.

2.2.2 Long Short-Term Memory (LSTM)

LSTM is a type of neural network that is capable of learning an order dependence in sequence order prediction. LSTMs are quite difficult to get hands around and are complex area of deep learning [16]. The application scope of LSTM covers speech recognition, handwriting recognition, and machine translation [17, 18] etc.

2.3 Student's Engagement

Student engagement is the involvement of student throughout the environment of learning [5]. According to the Skinner and Pitzer, Student Engagement in classroom has been divided in four types:

- Behavioral engagement
- Emotional engagement
- Cognitive engagement
- Psychological engagement [6]

2.4 Traditional Solution

According to the Hobsons Mission Statement, the traditional solutions are to:

- Ask student about their opinion and act on that opinion
- Give them a chance to discover themselves
- Appreciate the student on their achievement and engage them socially [7]

2.5 Technological Solutions

2.5.1 Kinect

In the study of Janez Zaletelj and Andrej Košir, Kinect has been used for the automated detection of student attention in the classroom using Machine Learning. They took data set of the undergraduate class of the university. Moreover, they wrote an algorithm that was used to detect the students' attention in classroom.

Kinect features of automated attention estimation are from:

- Kinect signal preprocessing
- Inter-person normalization and smoothing
- Computed feature set for attention estimation
  - Upper body posture
  - Face gaze point
  - Facial features
- Classifiers for attention estimation
- Training datasets and data splitting [8]

2.5.2 Eye Tracker

According to Narayanan Veliyath et al., the learning process is not only determined by what the instructor teaches, but the main aspect is how student grasp that information. The student who is attentive will obtain more knowledge then the student who is bored and frustrated during a lecture [9].

3 Methodology

Two types of neural network have been used to detect the attention of students in the classroom. One is Convolutional Neural Network (CNN) and other is Long Short-Term Memory (LSTM) on the FER-2013 dataset. LSTM is a variant of neural networks that has proven its importance in the technological applications which require sequence learning, such as human action recognition, handwriting recognition and time series prediction [11].

3.1 Dataset

FER-2013 is state of the art dataset for facial emotions. Dataset includes 35888 raw images for both single and multi-face. 28710 images are used for training and 7178 images are used for validation. It has 7 emotions (Happy, Anger, Surprised, Neutral, Sad, Fear and Disgust) in it. We distribute the 7 emotions in 2 classes which are "Attentive" and "Non-Attentive". The emotions include in attentive class are Happy, Anger, Surprised and Neutral. On the other hand, the remaining emotions are included in the non-attentive class which are Sad, Fear and Disgust.

3.2 Attention Detection Using Convolutional Neural Network

In this neural network, we trained the model on 6 network layers with a dropout 0.2, activation function is "relu" while on the last layer we used "softmax" as an activation function, optimizer is "adam" and loss function is "categorical cross entropy". We have trained our model on 30 epochs with a batch size of 64.

3.3 Attention Detection Using Long Short-Term Memory

In the long short-term memory neural network, the model is based on 14 network layers with an activation function "relu" while on the last layer we used "softmax" as an activation function, optimizer is "adamax" and loss function is "categorical crossentropy". We have trained our model on 5 folds, each fold has 40 epochs with a 64-batch size.

4 Results

4.1 Attention Detection Using Convolutional Neural Network

After training of the model on FER-2013 dataset the training accuracy is 0.9825 and the validation accuracy is 0.7623 (Fig. 1).

4.2 Attention Detection Using Long Short-Term Memory

After training of the model on FER-2013 dataset the training accuracy is 0.9940 and the validation accuracy is 0.7136 (Figs. 2 and 3).

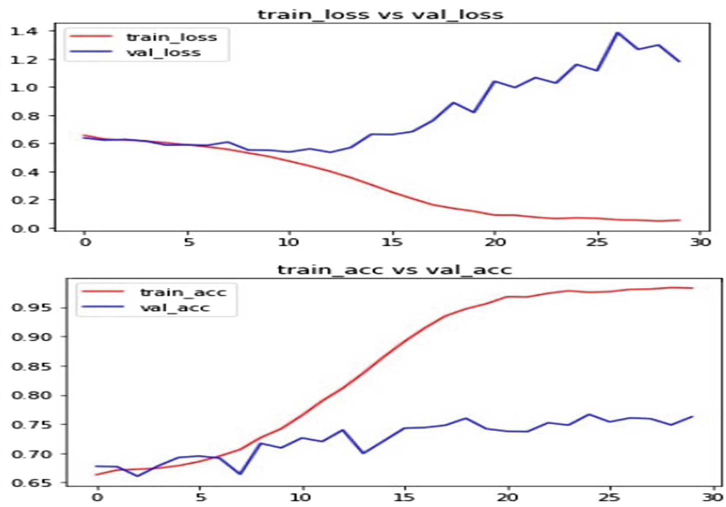

Fig. 1. Loss and accuracy of CNN

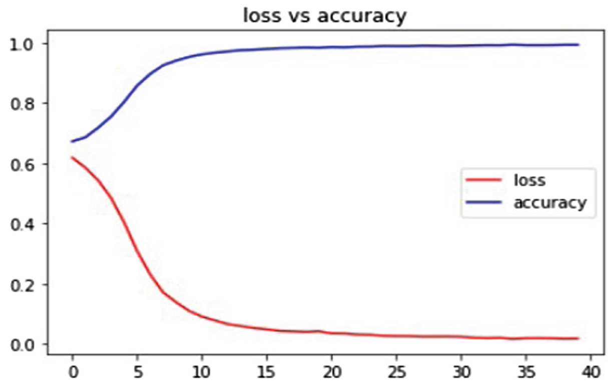

Fig. 2. Loss versus accuracy of LSTM

	precision	recall	f1-score	support
0	0.57	0.50	0.54	2349
1	0.77	0.82	0.79	4828
accuracy			0.71	7177
macro avg	0.67	0.66	0.66	7177
weighted avg	0.71	0.71	0.71	7177

Fig. 3. Precision chart of LSTM

4.3 Comparison of Results Between CNN and LSTM

After the compilation of results for both Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM), it can clearly be seen that the calculated results of CNN are much better than the LSTM.

For better results on this dataset FER-2013 through CNN and LSTM, we have increased and decreased the number of layers and have used different activation functions and loss functions. The above functions that we have discussed in methodology gave us better results for both CNN and LSTM.

5 Conclusion

The purpose of this study is to detect the attention of students in classroom based on the emotions. We have used FER-2013 dataset because we found it the best suitable dataset available for our study. Through this model, teachers can easily evaluate their teaching pedagogy and the attention of their students during the lecture. For getting better accuracy of this model, we can add more images in FER-2013 dataset or by creating a customized dataset for emotions with high definition image samples. For future researchers, it is suggested to carry out the same study in any kind of industry as the attentiveness and non-attentiveness of employees during work could help to identify the weaknesses and strengths of an organization.

Acknowledgement. We would like to acknowledge School of Electrical Engineering and Computer Sciences (SEECS), National University of Sciences and Technology (NUST), Islamabad, for technical support and funding.

References

1. Park, S.: Student engagement and classroom variables in improving mathematics achievement. Asia Pacific Educ. Rev. **6**, 87–97 (2005). https://doi.org/10.1007/BF03024970
2. Kraushaar, J.M., Novak, D.C.: Examining the affects of student multitasking with laptops during the lecture. J. Inf. Syst. Educ. **21**(2) (2019). Article 11
3. Nicole: Kaggle, 02 April 2019. https://www.kaggle.com/nicolejyt/facialexpressionrecognition
4. Krithika, L.B., Lakshmi Priya, G.G.: Student emotion recognition system (SERS) for e-learning improvement based on learner concentration metric. Procedia Comput. Sci. **85** (Cms), 767–776 (2016)
5. Martin, J., Torres, A.: User's Guide and Toolkit for the Surveys of Student Engagement: The High School Survey of Student Engagement (HSSSE) and the Middle Grades Survey of Student Engagement (MGSSE) I. What is Student Engagement and Why is it Important? (2016). http://www.isbe.net/learningsupports/pdfs/engagement-concept.pdf. Accessed 04 Oct 2019
6. Skinner, E.A., Pitzer, J.R.: Developmental dynamics of student engagement, coping, and everyday resilience. In: Christenson, S., Reschly, A., Wylie, C. (eds.) Handbook of Research on Student Engagement, pp. 21–44. Springer, Boston (2012). https://doi.org/10.1007/978-1-4614-2018-7_2

7. Hobsons.com: How to Increase Student Engagement at your School A Guide to Connecting Learning and Life Through Three Dimensions of Student Engagement 2 About Hobsons (2013)
8. Zaletelj, J., Košir, A.: Predicting students' attention in the classroom from Kinect facial and body features. Eurasip. J. Image Video Process. **2017**(1), 80 (2017)
9. Veliyath, N., De, P., Allen, A.A., Hodges, C.B., Mitra, A.: Modeling students' attention in the classroom using eyetrackers. In: ACMSE 2019 – Proceedings of the 2019 ACM Southeast Conference, pp. 2–9 (2019)
10. Salman, A., et al.: Real-time fish detection in complex backgrounds using probabilistic background modelling. Ecol. Inform. **51**, 44–51 (2019)
11. Afzal, M.Z., et al.: Document image binarization using LSTM: a sequence learning approach. In: Proceedings of the 3rd International Workshop on Historical Document Imaging and Processing (2015)
12. Siddiqui, S.A., et al.: Automatic fish species classification in underwater videos: exploiting pre-trained deep neural network models to compensate for limited labelled data. ICES J. Mar. Sci. **75**(1), 374–389 (2018)
13. Uzair, M., et al.: Representation learning with deep extreme learning machines for efficient image set classification. Neural Comput. Appl. **30**(4), 1211–1223 (2018)
14. Van den Oord, A., Dieleman, S., Schrauwen, B.: Deep content-based music recommendation. In: Advances in Neural Information Processing Systems (2013)
15. Collobert, R., Weston, J.: A unified architecture for natural language processing: deep neural networks with multitask learning. In: Proceedings of the 25th International Conference on Machine Learning (2008)
16. Huang, Y., Wang, W., Wang, L.: Instance-aware image and sentence matching with selective multimodal LSTM. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2017)
17. Sak, H., Senior, A., Beaufays, F.: Long short-term memory recurrent neural network architectures for large scale acoustic modeling. In: Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, pp. 338–342 (2014)
18. Graves, A., Liwicki, M., Fernández, S., Bertolami, R., Bunke, H., Schmidhuber, J.: A novel connectionist system for unconstrained handwriting recognition. IEEE Trans. Pattern Anal. Mach. Intell. **31**(5), 855–868 (2009). https://doi.org/10.1109/TPAMI.2008.137