Optimization The Utility Of E-Learning Platform Through Integrating Smart Emotional Recognition Feature

Layth Mohammed Abbas Al- Mashhadani[1], Isam Abdulmunem Alobaidi[1] and Alaa Mohammed Abbas Al- Mashhadani[1]

[1] Al-Farabi University College, Dept. of Computer Engineering, Baghdad, Iraq

E-mail: layth.muhammad@alfarabiuc.edu.iq

Abstract

Educational applications of image processing have emerged due to data collection tools development. Education is vital field in human life where highly accurate performance is required. Integrating image processing and deep learning with the education will help to optimize the performance of entire system. It is possible now to make out the student's emotional status through study the features from facial images taken for a group of students. That reduces the time and cost of the education by providing a facility similar to the regular classrooms environments. Which may help plenty of people who are unable to access regular educational facilities due to intolerable cost. In this paper, automatic emotional detection is being performed using neural network. Two models are used namely artificial neural network and CNN neural network. The models are tested using emotional images data. Results are reported 96.7 % and 99.2 % accuracies from bother artificial neural network and CNN respectively.

Keywords: Classrooms, LMS, E-learning, Voice, Features, Classification.

1. Introduction

Education is crucial for the survival of humanity and is valued on par with other necessities of life (i.e. food). However, in pandemic scenarios, the education sector is hit by severe lockdowns that interfere with the operations of institutions and schools. Virtual classroom technology is being used as a solution to deal with the issue. Following are a number of difficulties, particularly those relating to the distance between teachers and students. To connect with students effectively during lessons, instructors must be aware of their emotional state.

In the human-machine interaction, emotional recognition is a crucial area where computers can determine people' emotional states. [1]. This technology have wide spread in various applications including health, security, marketing, human resources, call centers, etc. more recently, learning management systems (LMS) have intensively utilized to tackle educational obstacles in events where students and tutors are unable to reach the colleges/schools [2]. Finding an effective and dependable replacement for the conventional educational systems was the motivation behind the development of LMS. The most likely beneficiaries of this technology are pupils who live in rural places and have significant difficulties getting to school on time. The invention of a virtual education portal for knowledge transmission was spurred on, however, by other issues such as financial crises when students cannot afford the enrollment fees or pandemic scenarios where educational organizations themselves are not operating. Although it outperformed during crises events, the virtual classes are utilized for supporting education quality alongside with conventional classrooms [3]. Finding an effective and dependable replacement for the conventional educational systems was the motivation behind the development of LMS. The most likely beneficiaries of this technology are pupils who live in rural places and have significant difficulties getting to school on time. The invention of a virtual education portal for knowledge transmission was spurred on, however, by other issues such as financial crises when students cannot afford the enrollment fees or pandemic scenarios where educational organizations themselves are not operating [4]. Furthermore, educational enhancement experience is also supported in [5] and [6] by incorporating of graphical objects e.g. avatars form motivating students/learners to enroll in education. Gamification of learning system is proposed at [7] for facilitate knowledge delivery. Other attempts were made for enhancement of educational experience through using of biometrical recognition such as [8] [9] [10], brain electrical activity using electroencephalogram (EEG) [11], pattern recognition for physical activity detection [12], [3]. Eye contact between the teacher and students is crucial for conveying information because it shows how open the student is to understanding the material being covered in class. Electronic learning systems (e-learning) may involve tens of learner at the time (number may be more), however, it is challenging for the tutor to follow the eye contact, face emotional actions. Thus, learning about alternative methods for performing of emotional recognition is important for e-learning platforms. Recognizing of the learners emotional status is curtail in as it helps the tutor to turn the lecture topics as per the learner readiness to receive the knowledge. Due to their outstanding role in e-learning systems efficiency enhancement, emotional recognition systems (ERS) were been a focal of many research activities. In this paper, survey of the most proposed state of the arts in emotional recognition and behaviors recognition is made.

2. Rescores and structure

ERS is establishing of connection between different emotional status and their respective features. Six emotional status are widely adopted in recognition process namely anger, fear, sadness, happiness, disgust and surprise.

ERS utilizes different sources of data such as voice, images, video and statistics while performing of emotional recognition [9]. Automatic ERSs are more advanced and reliable breeds of the system which are employing of deep learning and machine learning algorithms for performing of emotional recognition. Images are most popular source of emotional recognition followed by the voice. Large number of features can be extracted from an image in regard of emotional status. More specifically, face image can be utilized for same since face can reveal more than any other organ about human emotions. However, image is basically two dimensional (2D) data that contains of rows and colours of pixels where each pixel is represented by a number. The variance among pixels values are forming the information of an image as it seen by naked eye. Extracting of features corresponding with emotional status or any other biometrical information from an image is challenging task due to the none linearity, none stationary nature of image data [4]. That is manifested by features changing due to physical or environmental influences. More specifically, same features is impossible to be extracted from same person image within particular amount of time. Aging [5], is one of the most challenges in image recognition system (IRS). Other disturbances alike bad illumination [6], face wear and face orientation [7], noise due to dusts [8] are the common degradations of ERS performance. From the other hand, voices are crucial raw materials in ERSs process and as well were widely used in by previous researches. Activity recognition is another type of important emotional detection approaches [3]. This technology utilized images data in order to extract the emotions according to person limbs movement. Image processing technologies are valid for extracting those features.

Brain activity is one of outstanding approaches that can be utilized in various applications including emotional detection. EEG data is used for this purpose in many previous researches such as [4] and [5]. this data can be recorded using set of electrodes that to be placed over the skull. EEG capturing may be taking place either by invasive or none-invasive ways where the first is performed by placing Nano electrodes underneath of skull through a surgical procedure. The popular method of EEG signal capturing is termed as international 10-20 electrodes [4]. The signal capturing is performed in multiple channel from different regions of the brain where the emotional recognition can be achieved in corresponding with location and voltage level of the captured signal. The conversion of EEG electrical form into data is mainly performed using ADC same as voice signals. In all of the aforementioned sources of emotional data, pre-processing is vital for successful recognition process. Noise and other interferences can be eliminated from the said data by commonly used wavelet transform method [6].

With development of machine learning algorithms, new approaches had implemented to provide supper learning capability and enhanced prediction performance. Neural network is inspired by the human neural system and how does the neuron cells are working [7-9]. It reflects the same fundamentals of human learning system. The term artificial is allotted to those kinds of algorithm which represents smart generation of machine learning technology. Basically, neural network works in two stages namely: leaning stage and testing stage.

During the learning stage, neural network study the data and target and reach to concept that relating the both data at the input and the results (or the expected results). The following stapes can be considered while training stage.

Neural network is basically constructed of three parts namely input layer, hidden layer and output layer. The number of hidden layers are always differs from single to multiple layers depending of the design requirements. Figure 1 demonstrates the construction of single hidden layer neural network.

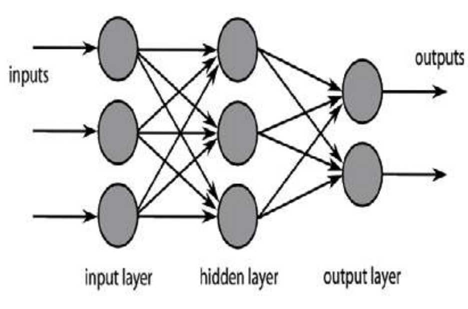

Fig 1. Neural network fundamental structure.

Each layer in the neural network is constructed with several (multiple) nodes, nodes number is decided according to the input length and generally, large node number might increase the execution time of the network. Nodes are connected with each other by means of weights, the weights are acting alike neurons in human nerves system. These weights are represented by numerical value which represents the number that scale the input when the input passes from a layer to another layer. Let x[n] is single dimensional array of one column and n rows which is passed into the neural network model at the input layer. The output of the hidden layer is expressed by y[n] and can be written as:

$$y[n] = \sum_{n=1}^{n=N} x[n].w_i \tag{1}$$

Whereas, the z[n] can stand for the output of the hidden layer and can be expressed as in the following equation.

$$z[n] = \sum_{n=1}^{n=N} y[n].w_h \tag{2}$$

It is worth to say that three different weight coefficients to be presented on the network namely input layer weight, hidden layer weight and output layer weight. However, the final output of the network model can be expressed as in equation below which entitled as m[n].

$$m[n] = \sum_{n=1}^{n=N} z[n].w_o \tag{3}$$

Training stage: neural network is firstly trained on specific data and training stage aims provide the network with all information related to the data structure and nature of this data more likely, how data elements are related to each other and how do they relate to the target. This process is essential task in the neural network operation. The aim accuracy of the neural network is depending on how the weight allotment is accurate.

Weight allotment is basically performed to satisfy the equation (3). Weight is the main role player in the process of output generation or in other word, the process of mapping the input to its particular class. Let the input to be x[n], the y[n] is to be the output from this process and the weight general formula is represented by W and the output of after passing form the weight is represented by following equation:

$$W = \frac{x[n]}{y[n]} \tag{4}$$

Knowing that weight vector can contain large number of elements which might be reaching thousands of weights which always depends on the number of dataset elements. Weight is being allotted to the neural network model using optimization algorithms more likely LM algorithm which acts as standard weight generation algorithm in neural network model.

3. CNN

Alex Krizhevesky has a remarkable academic career, with his 2012 ImageNet competition triumph being the crowning achievement (Krizhevsky, A., Sutskever, I., & Hinton, G. E. 2012). AlexNet is leading the way in the field of computer vision, excelling in both machine learning and conventional computer programming approaches. Important for visual recognition and classification in machine learning and computer vision, Deep learning soon grew, and in the history of increasing attention in computer vision. Figure 2 illustrates the architecture of AlexNet (Figure 2). A pool of 96 unique layers emerges from the first transformation and local response standardization. 11 * 11 squares are used to receive the signal. The three-by-three filters use two different stride measures, one of which is of one size and the other is larger. The identical second-layer filtering processes are completed with 5 * 5 filters. Filters. The filter types utilized in the third, fourth, and fifth convolution layers have a map of 384, 384, and 296, respectively, and consist of three and three filters, respectively. After the dropout, the network has two completely linked layers that finish the softmax layer. Two networks, which have the same structure and characteristic maps, comprise this model. LRN and dropout have been implemented in this network. LRN is implemented in two ways. Firstly, when a single channel is normalized using the neighborhood values, either the N to N patch or feature maps are appropriate. Second, the LRN can be extended by character sets or keyboard and on-screen displays (a third dimension neighborhood with just one pixel or location).

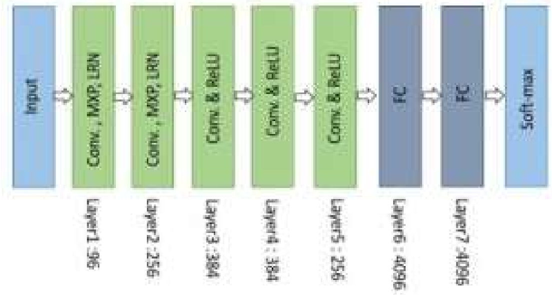

Fig 2. AlexNet's architecture uses convolution, max-pooling, learning rate-normalization, and completely connected (FC) layers.

AlexNet features three convolution layers, two of which are completely connected. Averaging the total amount of parameters at each layer for AlexNet reveals a starting point of 17.7K parameters when handling the ImageNet dataset. The input is 224×224×3 and the output is 55×55×96, with a stride of 4. The first layer's weight was calculated as 290400 neurons (with 55 total) and 364 weights, as shown in Section 3.1.4. Of the 105,705,600 parameters of the first convolution layer, they are 290400×364= 105,705,600 of them. In Table II, millions of parameters are shown each sheet. The network has a total weight of 61 million and MACs of 724 M.

Table 1. Accuracy measure of the proposed models.

Model	Accuracy
ANN	96.7
CNN	99.2

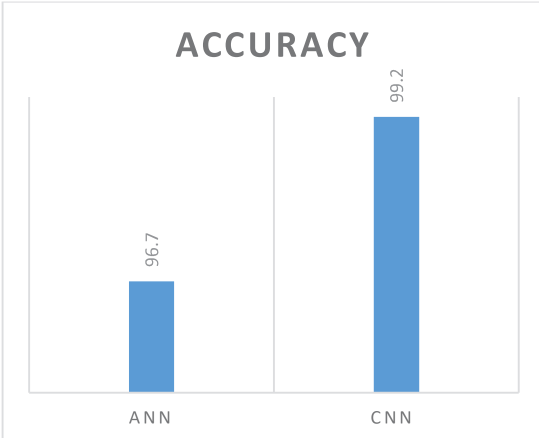

Fig 3. A graphical demonstration of Accuracy measure of the proposed models.

Table 2. Mean square error measures of the proposed models.

Model	MSE
ANN	0.00164
CNN	0.00091

Fig 4. A graphical demonstration of Mean square error measures of the proposed models.

4. Conclusions

Education is crucial for the survival of humanity and is valued on par with other necessities of life (i.e. food). However, in pandemic scenarios, the education sector is hit by severe lockdowns that interfere with the operations of institutions and schools. Virtual classroom technology is being used as a solution to deal with the issue. Following are numerous difficulties, particularly those relating to the distance between teachers and students. To connect with students effectively during lessons, instructors must be aware of their emotional state. The raw data for emotional recognition was examined from four main sources: facial images, voice signal, brain activity (EEG), and individual activity (e.g. limbs movement). Overall, all of the systems under evaluation are designed to categorize emotions into six classifications (anger, fear, sadness, happiness, disgust and surprise). Deep learning and machine learning can be combined (hybrid) or used separately to create ERS. Data comes in a variety of formats and dimensions, which makes it difficult for deep learning classifiers to handle it. The classifier configurations were changed by the researchers to increase classification accuracy by adding additional filters or layers. Filtering and noise reduction techniques may offer more payload to the ERS when dealing with other data sources, such as speech and EEG, for emotional detection purposes. Convolutional neural network (CNN) and artificial neural network are the two models that are utilized (ANN). After the models are trained, the accuracy and mean square error are determined as performance metrics (MSE). The CNN (AlexNet) neural network is discovered to have the lowest loss mean square error and the maximum accuracy of 99.2%.

References

[1] Fitzmaurice, C., Allen, C., Barber, R. M., Barregard, L., Bhutta, Z. A., Brenner, H., & Dicker, D. J. (2017). A systematic analysis for the global burden of disease study. JAMA Oncol, 3(4), 524-548.

[2] Fletcher, C. D., Unni, K., & Mertens, F. (2002). World Health Organization classification of tumours. Pathology and genetics of tumours of soft tissue and bone. IARC press.

[3] Motlagh, M. H., Jannesari, M., Aboulkheyr, H., Khosravi, P., Elemento, O., Totonchi, M., & Hajirasouliha, I. (2018). Breast cancer histopathological image classification: A deep learning approach. BioRxiv, 242818.

[4] Gurcan, M. N., Boucheron, L. E., Can, A., Madabhushi, A., Rajpoot, N. M., & Yener, B. (2009). Histopathological image analysis: A review. IEEE reviews in biomedical engineering, 2, 147-171.

[5] Rahman, A., Lee, J., & Choi, K. (2016, March). Efficient FPGA acceleration of convolutional neural networks using logical-3D compute array. In 2016 Design, Automation & Test in Europe Conference & Exhibition (DATE) (pp. 1393-1398). IEEE.

[6] Venieris, S. I., & Bouganis, C. S. (2016, May). fpgaConvNet: A framework for mapping convolutional neural networks on FPGAs. In 2016 IEEE 24th Annual International Symposium on Field-Programmable Custom Computing Machines (FCCM) (pp. 40-47). IEEE.

[7] Irshad, H., Veillard, A., Roux, L., & Racoceanu, D. (2013). Methods for nuclei detection, segmentation, and classification in digital histopathology: a review—current status and future potential. IEEE reviews in biomedical engineering, 7, 97-114.

[8] Filipczuk, P., Fevens, T., Krzyżak, A., & Monczak, R. (2013). Computer-aided breast cancer diagnosis based on the analysis of cytological images of fine needle biopsies. IEEE transactions on medical imaging, 32(12), 2169-2178.

[9] George, Y. M., Zayed, H. H., Roushdy, M. I., & Elbagoury, B. M. (2013). Remote computer-aided breast cancer detection and diagnosis system based on cytological images. IEEE Systems Journal, 8(3), 949-964.

[10] Spanhol, F. A., Oliveira, L. S., Petitjean, C., & Heutte, L. (2015). A dataset for breast cancer histopathological image classification. Ieee transactions on biomedical engineering, 63(7), 1455-1462.

[11] Spanhol, F. A., Oliveira, L. S., Petitjean, C., & Heutte, L. (2016, July). Breast cancer histopathological image classification using convolutional neural networks. In 2016 international joint conference on neural networks (IJCNN) (pp. 2560-2567). IEEE.

[12] Akbar, S., Peikari, M., Salama, S., Nofech-Mozes, S., & Martel, A. (2017). Transitioning between convolutional and fully connected layers in neural networks. In Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support (pp. 143-150). Springer, Cham.