Literature Review on Human Behavioural Analysis Using Deep Learning Algorithm

R. Poorni, P. Madhavan

Literature review on human behavioural analysis using deep learning algorithm

Mrs. R. Poorni[1], Dr. P. Madhavan[2]

[1]Department of Computer Science and Engineering, Easwari Engineering College, Ramapuram, Chennai, poorniram21@gmail.com.

[2]Department of Computing Technologies, SRM Institute of Science and Technology, kattankulathur, Chennai, madhavap@srmist.edu.in.

ABSTRACT:

Human behaviour analysis is the active area of research in computer science and engineering which determines the behaviour of humans using various algorithms. The input can be taken from the real time environment to analyze the human predictions. Deep learning plays a vital role as the input data involves a lot of computational images and spatial and temporal information upon which the predictions can be made. In this paper, we discuss the various techniques, concepts and algorithms that are implemented on a various field of image analysis and on real world input data to visualize the behaviour of a human.

Keywords: Real time environment, Deep learning algorithms, Behaviour analysis.

INTRODUCTION:

In today's digital world, the factual information that we get from our day to day is getting massive. In the current trend, a statistics says that the data produced and used will reach around 175 ZettaBytes in the year 2025. This increasing amount of data leads to the evolution of various new fields such as data mining, data science, artificial intelligence and its subset areas, etc.,. The major impulse of artificial intelligence is the data. The data is fed into the AI system to learn, analyze and perform the tasks accordingly. The fact that the amount of data to be handled becomes tremendous and more the number of interpretations that can be done with the data which lets the system to learn by itself that leads to machine learning. By the use of data the algorithm learns by itself and improves the predictions or decision making from the experience gained. A model is built based on training data or sample data to make decisions without any external interventions. Machine learning is closely intertwined with statistical computing as the majority of the decisions made are based on the probability of their occurrence. It uses the neural network to duplicate the human brain to perform a task without being programmed by the user. The model can be trained with the training set with the desired output it produces for the corresponding input which makes it supervised learning. Whereas unsupervised learning is the model created to work on the prediction without providing it with the structured data. The model learns by itself to

identify the structure of the input fed into it and produce the result. When the model works over the dynamic environment with changing input and output reinforcement model is implemented. Various machine learning algorithms are used to develop applications based on the above models for decision making.

When the data used and processing gets huge, deep learning which is the subset of machine learning is used. The rising level of data in the current era will give more convenience to explore the models that can be implemented using deep learning. It helps to solve complex problems with more detailed representation of images with the help of neural networks. Neural network acts as a heart of deep learning algorithms as it contains more numbers of hidden layers into it. Each layer acts as a processing unit in a level of hierarchy. That is, the first layer will compute for a lower level of element from the image or an object. The result of this first layer is fed into the second layer as an input and in the second layer the next level of element is extracted and the output is fed into the next layer and so on until the desired output is achieved. The depth or the number of layers of the deep learning algorithms are considered based on Credit Assignment Path (CAP). It analyzes how to choose the data points so that the error or backtracking gets minimal. Each hidden layer is given with the weight to the input which will be adjusted based on the corresponding result obtained and the accuracy of the output. Since there is a huge amount of data and there are a wide range of parameters included, the deep learning model takes time to get trained. But testing the model will take less time. The output of every node is activated by the activation function which helps to remove the unwanted data from the useful information. The various research work done under the domain are discussed.

RELATED WORK:

With the help of deep learning, image processing, face identification and analysis and Neural Network algorithm, many earlier studies on human behaviour analysis have been done. In this section, some recent research work done by various authors are discussed.

[1] Nilay Tufek, Murat Yalcin, Mucahit Altintas, Fatma Kalaoglu, Yi Li, and Senem Kursun Bahadir conducted research on “Human Action Recognition Using Deep Learning Methods on Limited Sensory Data”. Here, the accelerometer and the gyroscope data is used in a lesser amount to build an action recognition system. The Convolutional Neural Network, Long-Short Term Memory (LSTM) and various combinatorial algorithms were used and their performances were compared. The accuracy of the model was increased by augmenting the data. By using a 3 layer LSTM model, the accuracy of the model was increased by 97.4%. Over the collected data set the system produced 99% of accuracy. To analyse the performance precision, recall and f1-score metrics were also used. For the purpose of evaluation on the classification an application is developed using a 3 layer LSTM network.

To detect human activity a unique set of wearable devices were used. When the dataset is high, the 3 layer LSTM model provides high accuracy to sensory data when compared to KNN model. To improve the test accuracy, precision, recall and F1 Score radically the data augmentation is done on a small size dataset. The time series sensory data plays a vital role for the implementation of the system. To analyze and recognize more intricate behaviour, multiple sensor data can be used.

[2] Kai Zhang and Wenjie Ling proposed a research work on “Joint Motion Information Extraction and Human Behavior Recognition in Video Based on Deep Learning”. Here it mainly focuses on human behaviour analysis and identification from videos. A two channel deep convolutional neural network model is designed for the structure and joint motion feature extraction. The network structure also simulates the brain visual cortex which processes the visual signal. The static information is processed by the spatial channel network and the dynamic information is processed by the temporal channel network. The spatial and the temporal information are extracted separately. The recognition of the two channel model is compared with the single channel model to verify the superiority of the dual channel structure. KTH behavioural dataset are used to perform the experiments and shows that the human behaviour analysis using deep learning has achieved more accuracy based on joint motion information. The joint motion feature extraction using deep convolutional neural networks removes the extra calculations and multifaceted computations that are involved in conventional feature extraction methods. This model can be applied to a wide variety of video data using which a more detailed understanding of action behaviour can be done.

[3] Dersu Giritlioglu, Burak Mandira, Selim Firat Yilmaz, Can Ufuk Ertenli, Berhan Faruk Akgur, Merve Kiniklioglu, Aslı Gül Kurt, Emre Mutlu, Seref Can Gurel, Hamdi Dibeklioglu proposed a paper on “Multimodal analysis of personality traits on videos of self-presentation and induced behavior”. In this paper a multimodal deep architecture is presented to estimate the personality traits from audio-visual cues and transcribed speech. For the detailed analysis of personality traits audio visual dataset with self presentation along with recordings of induced behaviour is used. The reliability of various behaviours and their combinational use is assessed. The face normalisation is done by marking 68 boundaries of the face region from the video using OpenFace. The frontal view of the face is obtained by performing translation, rotation and scaling. The X and Y coordinates are identified and are shape normalized by using linear warping. Finally a facial boundary region with the resolution of 224*224 pixels are obtained. The normalized facial images are modelled with the spatio temporal pattern using ResNext model and CNN-GRU. The facial appearance and facial dynamics are captured from the input video. To access the temporal data ResNext model is used. The feature map is divided into smaller groups from the ResNet. The random

temporal sampling of 1.5 seconds is used during training and the same size data is used during the validation phase. The second layer of the deep learning architecture is modelled with CNN-GRU for facial videos. For the modelling, the action unit and the head pose uses LSTM and to obtain the gaze feature the model uses RCNN.The body pose is obtained with the 2D coordinates of the traced points. The audio of the video is extracted using pyAudio analysis framework. The speech language is automatically identified by the Google speech application programming interface. The testing performed over the SIAP and FID dataset. The openness and Neuroticism provides the social desirability effect. Based on the mean score, the participants give the rating of openness by 39.7% and neuroticism by 21.2%. For personality analysis, the face related models appear to be more reliable.

[4] Karnati Mohan , Ayan Seal , Ondrej Krejcar , and Anis Yazidi proposed a paper on “Facial Expression Recognition Using Local Gravitational Force Descriptor-Based Deep Convolution Neural Networks”. The proposed work contains two methods. Identification of local features from the image of faces with the help of local gravitational force descriptor and embedding the descriptor into the traditional deep convolutional neural network model which involves exploring geometric features and holistic features. The final classification score is calculated using the score level fusion technique. The pixel value of an image is considered to be the mass of the body. The gravitational force of the pixel is considered to the centre pixel on its adjacent pixel. A pixel is selected along with its adjacent pixels as the GF of the image. The DCNN learns by itself with the help of back propagation with more accuracy. The methods such as VGG-16 and VGG-19 are developed based on the single branch convolutional layers connected sequentially. No edge information is provided as it focuses only on the receptive fields and so there is not enough information on spatial structure of the face. This problem is addressed using multi convolutional networks. The first part of the architecture extracts the local features of the object from the image and the second part extracts the holistic features. The former part contains three convolutional layers that are in order as max pooling, average pooling and zero padding. The latter part consists of five convolutional layers which are merged and sent to the layers for classifying facial expression.This can extract the features automatically. Each layer is embedded with the filter and the bias value which is then fed into the activation function. The overfitting problem is avoided by applying max pooling to the feature maps obtained. The gradients of the facial images are trained independently and the probability of each layer is calculated. The final prediction of the basic expression is done by score level fusion. The model is trained using keras framework. The data is augmented to improve the performance.

[5] Yeong-Hyeon Byeon, Dohyung Kim, Jaeyeon Lee and Keun-Chang Kwak proposed a paper on “Explaining the Unique Behavioral Characteristics of Elderly and Adults Based on Deep Learning”. In this paper, it analyzes the behaviour of the elderly people

depending on their physical condition. The silver robots provide customized services to those who are in need to improve the betterment of the elders. When there is a change in the human body it gets reflected as pose evolution images. The classification is done based on convolutional neural networks which is trained based on the elderly behaviour. The gradient weighted activation map is obtained for the result obtained after classification and a heatmap is generated for the purpose of analysis. The skeleton heatmap and RGB video matching are analyzed and the characteristics are derived from this. To efficiently store the movement of the human based on reconstruction of the skeleton with coordinate points based on sensor data, the skeleton data type is used. The human body is modelled using kinect v2 with 25 joints. This also includes the head, arms, hips and legs. A sequence of skeletons is created like a video to get more information about the behaviour of a person. For the effective analysis of the skeleton sequence data both the temporal and the spatial information is used and to extract the appropriate information conversion methods are used. The skeleton sequence is converted into a color image called PEI method which is less readable for humans. The two dimensional convolutional neural network is used to extract the spatial and the temporal information from the skeletal sequenced image. The skeleton sequence is captured at the continuous time interval which has a 3D data type. Project the 3D coordinates into RGB space to convert the 3D image into 2D model. The product of number of joints, dimensionality of the coordinates and number of skeleton frames over time gives the skeleton sequence. The skeleton sequence can be converted into image by denoting the joint coordinate dimension with temporal dimension. The action analysis is conducted over the dataset of NTU RGB+D which has over 114480 data samples. The study was conducted over the average age of 77 years who performed 55 behaviours in their day to day life. After training the model identified the input data belongs to the elderly behaviour or young adults. For a more distinct behaviour the accuracy obtained was higher when compared to the normal behaviour.

[6] Angelina Lu, Marek Perkowski, proposed a research work on “Deep Learning Approach for Screening Autism Spectrum Disorder in Children with Facial Images and Analysis of Ethnoracial Factors in Model Development and Application“. Here, a Autism Spectrum Disorder screening solution is developed with the help of facial images by using VGG-16 transfer learning based deep learning technique. The kaggle ASD facial image dataset is used to implement the model. The resulting model can identify the children with ASD and normally developed children. The deep learning model of Visual Geometry Group VGG 16 has produced the classification with high accuracy. The reusing of model is achieved with the help of transfer learning and gives the output of one task to another. The model majorly focuses on the quality of detecting ASD in children with the help of facial images and the race factor of the children involved in the experiments. The accuracy and the

F1 score achieved are 95% and 0.95 respectively. Computer vision with facial images can be applied in screening the ASD children and further can be implemented to a user-friendly mobile application which helps the people in case of inaccessible medical emergencies.

COMPARISON:

TITLE	**INPUTS USED**	**ALGORITHMS/ METHODS USED**	**ACCURACY**
Human Action Recognition Using Deep Learning Methods on Limited Sensory Data	Accelerometer and the gyroscope data	Convolutional Neural Network, Long-Short Term Memory (LSTM) and various combinatorial algorithms	3 layer LSTM model - 97.4% OVERALL PERFORMANCE - 99% of accuracy
Joint Motion Information Extraction and Human Behaviour Recognition in Video Based on Deep Learning	Human behaviour analysis and identification from videos	Two channel deep convolutional neural network model	Joint motion information – 97%
Multimodal analysis of personality traits on videos of self-presentation and induced behaviour	Audio visual dataset with self presentation along with recordings of induced behaviour is used	Multimodal deep architecture	Openness - 39.7% neuroticism - 21.2% Face related models – 98%
Facial Expression Recognition Using Local Gravitational Force Descriptor-Based Deep Convolution Neural Networks	Image of faces - pixel value of an image is considered to be the mass of the body	Facial Expression Recognition Using Local Gravitational Force Descriptor-Based Deep Convolution Neural Networks	Score level fusion – 89%

Explaining the Unique Behavioural Characteristics of Elderly and Adults Based on Deep Learning	Elderly people behavioural video	Skeleton heatmap and RGB video matching	Conducted over the average age of 77 years who performed 55 behaviours – 93% of normal behaviours were classified correctly
Deep Learning Approach for Screening Autism Spectrum Disorder in Children with Facial Images and Analysis of Ethnoracial Factors in Model Development and Application	Kaggle ASD facial image dataset	Deep learning model of Visual Geometry Group VGG 16	The accuracy and the F1 score achieved are 95% and 0.95 respectively

DATASETS:

The datasets for the human behavioural analysis to analyse the human behaviour on various scenarios can be obtained by using data repository of Mendeley which is a free open source data repository. It contains 11 million datasets that are indexed for easy accessibility and contains various research data such as raw or processed data, videos, images, audios etc.,. Also BARD dataset can be used which is a collection of videos majorly used for the human behavioural analysis and detection. The major advantage of BARD video is that the dataset videos contain open environment captured videos and are collected in uncontrolled scenarios which will be helpful to implement in any real time applications. Various other open source datasets are available to make the study on behavioural analysis both in the controlled or uncontrolled environment.

CONCLUSION AND FUTURE WORK:

The model studies the various algorithms and techniques involved in analysing human behaviour using deep learning algorithms. While performing the study over various research papers it is found that the common algorithm that were used over the human analysis or the behavioural identification includes Convolutional Neural Network. Upon which various

models were built to improvise the performance or the accuracy of the classification. It is also identified that the data set used for the classification or the identification plays a significant role. When the data used for training or testing the model becomes higher the accuracy that the model produces is high. Thus the performance is directly proportional to the number of data used for training.

As a future work from this survey, a deep neural network can be implemented with an architecture for training the model to identify, analyse and classify the behaviour of the human is designed to produce high performance and accuracy.

REFERENCES:

[1] Nilay Tufek, Murat Yalcin, Mucahit Altintas, Fatma Kalaoglu, Yi Li, and Senem Kursun Bahadir, “Human Action Recognition Using Deep Learning Methods on Limited Sensory Data”, IEEE SENSORS JOURNAL, VOL. 20, NO. 6, MARCH 15, 2020.

[2] Kai Zhang and Wenjie Ling, “Joint Motion Information Extraction and Human Behavior Recognition in Video Based on Deep Learning”, IEEE SENSORS JOURNAL, VOL. 20, NO. 20, OCTOBER 15, 2020.

[3] Dersu Giritlioglu, Burak Mandira, Selim Firat Yilmaz, Can Ufuk Ertenli, Berhan Faruk Akgur, Merve Kiniklioglu, Aslı Gül Kurt, Emre Mutlu, Seref Can Gurel, Hamdi Dibeklioglu, “Multimodal analysis of personality traits on videos of self-presentation and induced behavior”, Journal on Multimodal User Interfaces (2021) 15:337–358 https://doi.org/10.1007/s12193-020-00347-7..

[4] Karnati Mohan , Ayan Seal , Ondrej Krejcar , and Anis Yazidi , “Facial Expression Recognition Using Local Gravitational Force Descriptor-Based Deep Convolution Neural Networks”, IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT, VOL. 70, 2021

[5] Yeong-Hyeon Byeon, Dohyung Kim, Jaeyeon Lee and Keun-Chang Kwak, “Explaining the Unique Behavioral Characteristics of Elderly and Adults Based on Deep Learning”, Appl. Sci. 2021, 11, 10979. https://doi.org/10.3390/app112210979

[6] Angelina Lu, Marek Perkowski, “Deep Learning Approach for Screening Autism Spectrum Disorder in Children with Facial Images and Analysis of Ethnoracial Factors in Model Development and Application“, Brain Sci. 2021, 11, 1446. https://doi.org/10.3390/brainsci11111446.

[7] Bala B., Kadurka R., Negasa G. (2022) Recognizing Unusual Activity with the Deep Learning Perspective in Crowd Segment. In: Kumar P., Obaid A.J., Cengiz K., Khanna A., Balas V.E. (eds) A Fusion of Artificial Intelligence and Internet of Things for Emerging Cyber Systems. Intelligent Systems Reference Library, vol 210. Springer, Cham. https://doi.org/10.1007/978-3-030-76653-5_9

[8] Benjamin Joseph Ricard, Saeed Hassanpour, “Deep Learning for Identification of Alcohol-Related Content on Social Media (Reddit and Twitter): Exploratory Analysis of Alcohol-Related Outcomes”, J Med Internet Res 2021 | vol. 23 | iss. 9 | e27314.

[9] MONAGI H. ALKINANI, WAZIR ZADA KHAN, QURATULAIN ARSHAD, “Detecting Human Driver Inattentive and Aggressive Driving Behavior Using Deep Learning: Recent Advances, Requirements and Open Challenges”, Digital Object Identifier 10.1109/ACCESS.2020.2999829.

[10] K. Prabhu1, S. SathishKumar, M. Sivachitra, S. Dineshkumar and P. Sathiyabama, “Facial Expression Recognition Using Enhanced Convolution Neural Network with Attention Mechanism”, Computer Systems Science & Engineering, DOI:10.32604/csse.2022.019749.

[11] Bhoomeshwar Bala, Raja Shekar Kadurka, and Galeta Negasa, “Recognizing Unusual Activity with the Deep Learning Perspective in Crowd Segment”, A Fusion of Artificial Intelligence and Internet of Things for Emerging Cyber Systems, Intelligent Systems Reference Library 210, https://doi.org/10.1007/978-3-030-76653-5_9.