Human Activities Recognition using different Machine Learning architectures

Under IIT Mandi - iHub & HCI Foundation Anubhav Fellowship

Submitted by

Sakshi Sharma, Shobhit Sharma, and Tushar Parmar

Jawaharlal Nehru Government Engineering College, Himachal Pradesh Technical University, Mandi, India – 175019

Under Supervision of

Dr. Varun Dutt

April 2022

Varun Dutt 10/4/22

Acknowledgements

We are pleased to acknowledge Dr. Varun Dutt for their invaluable guidance during the course of this project work; for his humble guidance, endless support, and understanding during the project duration. We extend our sincere thanks to Mr. Shashank Uttrani who continuously helped us throughout the project and without his guidance; this project would have been an uphill task. We are also grateful to other members of the ACS team who co-operated with us regarding some issues. The completion of this project could not have been possible without the participation and assistance of a lot of individuals contributing to this project. Last but not the least; we are also grateful to other coordinators of IIT Mandi - iHub & HCI Foundation who cooperated with us for the smooth development of this project.

April 2022

Sakshi Sharma

Shobhit Sharma

Tushar Paramar

Signature of

Prof. Varun Dutt

Associate Professor

School of Computing and Electrical Engineering

Indian Institute of Technology Mandi

Abstract

Previous research has looked at how humans succeed in basic psychological activities with less cognitive effort. However, little is known about how humans learn in sophisticated search-and-retrieve simulated scenarios. Our study's major purpose was to evaluate human performance in a challenging search-and-retrieve situation. We utilized Unity 3D to create a complex simulated environment that looked like a military on-the-ground operation, complete with targets and distractors. For a total of 25 minutes, 50 persons were asked to play the game simulation. The contestants' goal was to achieve the highest possible score. This can be accomplished by gathering targets and avoiding distractions. There were two phases to the game: testing and training. In contrast to the testing phase, the training phase provided feedback. The testing phase lasted ten minutes, whereas the training phase lasted fifteen minutes. Participants were free to explore the surroundings and collect items throughout the training period (7 distractors and 14 targets). According to the findings, human volunteers' performance altered dramatically across the training and test stages. The outcomes of the test phase without feedback were much better than those of the training phase with feedback. In addition, in both the train and test stages, the fraction of targets collected grew dramatically over time. Development of simulations for training humans in various jobs is one of the uses. In this, we used four neural network models, i.e., one RNN and three CNN models on the dataset that was obtained by playing the game in the Unity 3D environment.

Contents

Introduction

Machine learning is a subset of artificial intelligence that focuses on extracting patterns from large amounts of data that can improve on its own as a result of learning and data. Supervised learning, unsupervised learning, and reinforcement learning are the three major categories of machine learning algorithms.

This research focuses on Reinforcement Learning, which is a type of learning employs, a reward- penalty method to teach an AI system to learn from experience without the need for pre- programming and human supervision. It's a way for a computer programming to learn about a new environment and earn rewards by exploring and gaining success or failure experiences. Reinforcement learning's goal is to learn a policy, which is a mapping from observations to actions. This report discusses the result of the work done in development of “Cognitive algorithms and comparing them with already existing reinforcement learning algorithms" in Unity Environment. It is a part of the ASC (Applied Cognitive Science) project going in Computer Science Department, IIT Mandi and aims at the development of an NET framework for providing a common platform for facilitating the use of methodological approach developed by the ASC team and integration of various tools developed during the execution of the project.

Background and Motivation

Human activity recognition, or HAR, is a time series classification task that is difficult to master. It is the problem of classifying sequenced data that are recorded by smart phones into known well-defined movements. There are two traditional approaches for the time series-based data on fixed sized windows and training machine learning models, such as ensembles of decision trees.

It consists predicting a person's movement based on sensor data. It traditionally entails deep domain expertise and signal processing methods to correctly engineer features from raw data in order to fit a machine learning model. Deep learning methods such as convolutional neural networks and recurrent neural networks like LSTMs have recently demonstrated that they are capable of learning features from raw sensor data and can even achieve state-of-the-art results on challenging activity recognition tasks.

In this experiment, we'll learn about the problem of recognizing human activity and the recurrent neural network, and deep learning neural network models that are achieving state-of-the-art performance on it. The HAR project aims to analyze human behavior in a challenging search-and-retrieve scenario through artificial intelligence technology the notion is that given a system specification; the user can synthesize a system that fulfills his requirements by following the process and using the tools provided to support it.

Methodology

Participants

We recruited 50 participants from Indian Institute of Technology Mandi to perform in this experiment. Out of the total recruited participants 71% of them were male and the rest were female. The ages of the participants varied between 22 and 39 years of age. The mean age of the participants was 25.5 years with a standard deviation of 3.4 years. All the participants were renumerated with a fee of INR 50 for their participation in this experiment.

Experimental Design

A Unity 3D gaming environment was developed to simulate a search and retrieve game mission. There were four buildings in the simulation, each consist of targets and distractors that is present throughout the game. We randomly placed our targets and distractors in all directions all around the game. The players' objective was to get the highest possible score. This could have been accomplished by collecting the targets valued +5 points and distractor valued -5 points.

Our experiment consisted of two phases: training phase and test phase. During the training phase, 14 targets and 7 distractors were randomly distributed in the game and feedback was given to any of the participants. Whereas, during test phase, 28 targets and 14 distractors was disturbed with no feedback was given to any of the participants. Total time given to play a game is 25 minutes which was divided into 15 minutes and 10 minutes for training and test phase respectively. In order to generalize the learning environment various types of targets and distractors were developed in the testing phase.

Models

We are currently developing four different neural network models to predict human action based on the data collected in this experiment. Following are the four models:

Multilayer Perceptron (MLP)

A multilayer perceptron is a class of feedforward artificial neural network (ANN). The term MLP is used ambiguously, sometimes loosely to mean any feedforward ANN, sometimes strictly to refer to networks composed of multiple layers of perceptron's (with threshold activation). Multilayer perceptron are sometimes colloquially referred to as "vanilla" neural networks, especially when they have a single hidden layer [6].

An MLP consists of at least three layers of nodes: an input layer, a hidden layer and an output layer. Except for the input nodes, each node is a neuron that uses a nonlinear activation function. MLP utilizes a supervised learning technique called back propagation for training. Its multiple layers and non-linear activation distinguish MLP from a linear perceptron. It can distinguish data that is not linearly separable. A fully connected multi-layer neural network is called a Multilayer Perceptron (MLP). It has 3 layers including one hidden layer. If it has more than 1 hidden layer, it is called a deep ANN. An MLP is a typical example of a feedforward artificial neural network. In this figure, the ith activation unit in the ith layer is denoted as ai(l).

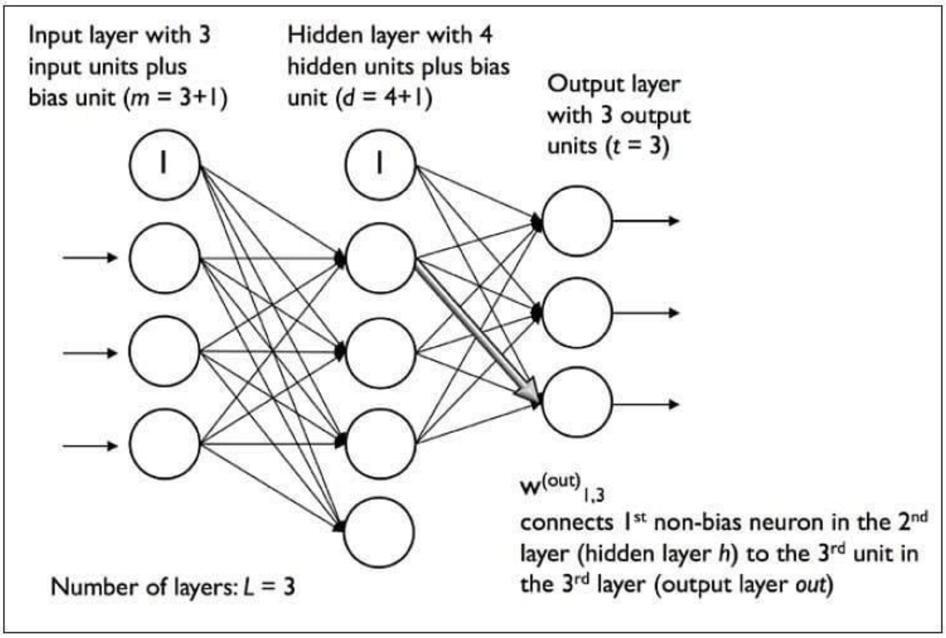

Figure 1. Multilayer Perceptron Network Architecture

The number of layers and the number of neurons is referred to as hyper parameters of a neural network, and these need tuning. Cross-validation techniques must be used to find ideal values for these. The weight adjustment training is done via backpropagation. Deeper neural networks are better at processing data. However, deeper layers can lead to vanishing gradient problems.

Long Short-term Memory (LSTM)

LSTM network models are a type of recurrent neural network that are able to learn and remember over long sequences of input data. They are intended for use with data that is comprised of long sequences of data, up to 200-to-400-time steps. They may be a good fit for this problem. The model can support multiple parallel sequences of input data, such as each axis of the accelerometer and gyroscope data. The model learns to extract features from sequences of observations and how to map the internal features to different activity types [7].

The benefit of using LSTMs for sequence classification is that they can learn from the raw time series data directly, and in turn do not require domain expertise to manually engineer input features. The model can learn an internal representation of the time series data and ideally achieve comparable performance to models fit on a version of the dataset with engineered features.

CNN Long Short-Term Memory Networks (CNN-LSTM)

Input with spatial structure, like images, cannot be modeled easily with the standard Vanilla LSTM. The CNN Long Short-Term Memory Network or CNN-LSTM for short is an LSTM architecture specifically designed for sequence prediction problems with spatial inputs, like images or videos.

The CNN-LSTM architecture involves using Convolutional Neural Network (CNN) layers for feature extraction on input data combined with LSTMs to support sequence prediction [8].

CNN-LSTMs were developed for visual time series prediction problems and the application of generating textual descriptions from sequences of images (e.g., videos).

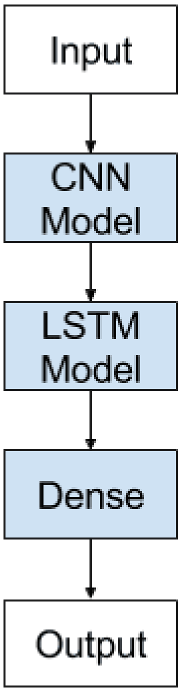

Figure 2. Convolutional Neural Network Long Short-Term Memory Network Architecture

Convolution Long Short-term Memory (Conv-LSTM)

In Conv-LSTM, Sequential images, one approach is using Conv-LSTM layers. It is a Recurrent layer, just like the LSTM, but internal matrix multiplications are exchanged with convolution operations. As a result, the data that flows through the Conv-LSTM cells keeps the input dimension (3D in our case) instead of being just a 1D vector with features [9].

CNN-LSTM is the combination of CNN layers and LSTM layers that ensures both advantages of CNN and LSTM. Fig. 4 shows the architecture of model based on CNN-LSTM for diabetes patients' classifications. In our application, we feed inputs into a convolution layer that extracts features from the given inputs and adds non-linearity in our convolutional network with the help of an activation function Rectified Linear Unit (ReLU). Then, the generated rectified feature maps are passed through a pooling layer that performs down sampling in order to decide the most activated features using max pooling operation. Then, the down sampled feature maps are passed through a tile of LSTM layers that handle the sequence processing. Lastly, a fully connected layer represents the feature vector and uses it for classification, regression with the help of SoftMax function.

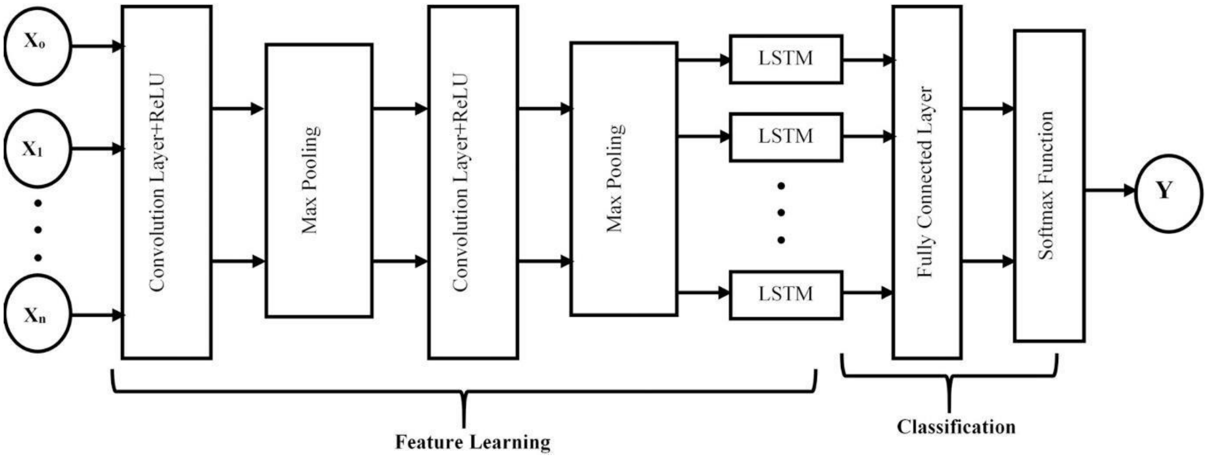

Figure 3. Architecture of Conv-LSTM model.

Procedure

A Unity 3D game environment simulation was developed as a search and retrieve simulation. For the maximization of the score, we recruited 50 participants from the IIT Mandi who played the game for the specific duration of the time. The total time of 25 minutes of the gameplay divided into two phases of training and test for 15 and 10 minutes respectively. We collected the data by playing the game in the form of video of training and test phase in which we took the screenshots frame by frame and reshaped in the vectorized form. Next, we applied principal component analysis (PCA) on the vector form of video data and choose 150 major components. Thus, the finally prepared data had position coordinates (x, y, z), reward, and PCA variables as independent variables and actions as the dependent variable. Finally, this dataset is being used for the development four machine learning models, i.e., MLP, LSTM, CNN-LSTM, and Conv-LSTM.

Discussion and Conclusion

In this study, it's feedback-based Machine learning techniques, such as Principal component analysis (PCA), are used to teach an agent how to behave in a given environment by performing actions and seeing the results of those actions. Principal Component Analysis is an algorithm that is used for the dimensionality reduction in machine learning. It is a statistical process that converts the observations of correlated features into a set of linearly uncorrelated features with the help of orthogonal transformation. It is one of the popular tools that are used for exploratory data analysis and predictive modeling. It is a technique to draw strong patterns from the given dataset by reducing the variances.

In this experiment we studied the problem of human activity reorganization and used one recurrent neural network approach, multilayer perceptron, and three deep learning method approaches, Long-Short-Term-Memory, CNN-LSTM, and Conv. LSTM. Both networks as well as the combination of two, are most suitable to extractions features from raw sensor data and predicting action. Neural Networks are forecasting methods that are based on predictions from the user data. It finds extensive applications in area where traditional computer doesn’t go far too well. Problem statement where system learns, adapt and changes the results on its own according to the data that we are providing it, rather using pre-programmed outputs. Let us take an example of chess players in the game utilize chess engines to analyze their games, improve their skills, and try new approaches- and it continues where it uses neural network to do so.

This research has prompted lots of new ideas that can be tested in the future. First, based on their diversity and level of required workload, similar simulations can be developed for different cognitive demand tasks. Second, computational cognitive models (such as instance-based learning models and machine learning models (such as Soft-Actor Critic and Proximal Policy Optimization can be developed to account for human choices in such tasks. Third, multi-agent simulations of such games can be created to see how humans perform in groups with other humans or robots in similar search-and-retrieve tasks. Furthermore, an LSTM-based Recurrent Neural Network with Time Series Classification for human activity recognition can improve the model's accuracy.

References

[1] J. Schulman, S. Levine, P. Moritz, M. I. Jordan, en P. Abbeel, “Trust Region Policy Optimization”, CoRR, vol abs/1502.05477, 2015.

[2] V. Mnih et al., “Asynchronous Methods for Deep Reinforcement Learning”, CoRR, vol abs/1602.01783, 2016.

[3] T. P. Lillicrap et al., “Continuous control with deep reinforcement learning”, arXiv [cs.LG]. 2019

[4] Schulman, John, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. "Proximal policy optimization algorithms." arXiv preprint arXiv:1707.06347 (2017).

[5] Haarnoja, Tuomas, Aurick Zhou, Pieter Abbeel, and Sergey Levine. "Soft actor-critic: Off- policy maximum entropy deep reinforcement learning with a stochastic actor." In International conference on machine learning, pp. 1861-1870. PMLR, 2018.

[6] An Overview on Multilayer Perceptron (MLP) [Updated]. [online]. Available at: https://www.simplilearn.com/tutorials/deep-learning-tutorial/multilayer-perceptron

[7] Brownlee, J. (2018). LSTMs for Human Activity Recognition Time Series Classification. [online] machinelearningmastery.com. Available at: https://machinelearningmastery.com/how-to-develop-rnn-models-for-human-activity-recognition-time-series-classification/ [Accessed 1 Jan. 1970].

[8] Brownlee, J. (2017). CNN Long Short-Term Memory Networks. [online] machinelearningmastery.com. Available at: https://machinelearningmastery.com/cnn-long-short-term-memory-networks/ [Accessed 1 Jan. 1970]

[9] Xavier, A. (2019). An introduction to ConvLSTM. [online] medium.com. Available at: https://medium.com/neuronio/an-introduction-to-convlstm-55c9025563a7 [Accessed 1 Jan. 1970].