Human Activity Recognition using Multilayer Perceptron Model and Convolutional Neural Networks

Vignesh Vasudevan (vasudevan.v@northeastern.edu)

Rajaragunanthan Palanisamy (p.r@northeastern.edu)

Sujai Rajan (rajan.suj@northeastern.edu)

Northeastern University, College of Engineering,

Data Analytics and Engineering Program, Boston, MA, USA

Abstract

Nowadays, Human Activity Recognition (HAR) is being employed in a range of domains, and vision and sensor-based data make use of up-to-date technologies to observe, recognize, and monitor human activities. many reviews and surveys on HAR have already been printed. The articles reviewed during this study are classified to spotlight application areas, knowledge sources, techniques, and open analysis challenges in HAR. The bulk of existing analysis seems to own targeting daily living activities, followed by user activities supported individual and group-based activities. However, there's very little literature on real-time activity detection of time-sensitive activities like suspicious activity, police work, and attention. A large portion of existing studies has used (CCTV) videos and Mobile Sensors data. Convolutional Neural Network (CNN), Long immediate memory (LSTM), and Support Vector Machine (SVM) constitute the foremost outstanding techniques within the research currently being utilized for the task of HAR. Lastly, the restrictions and open challenges that are required to be addressed have also been reviewed.

Introduction

In their daily lives, people engage in a diverse range of activities. Closed-circuit television (CCTV) and sensor data have recent advancements in technology, allowing for the detection of anomalies as well as the identification of regular human activity for surveillance. Anomalies are defined as aberrant or extraordinary behavior or activities. To recognize distinct human activities, Human Activity Recognition (HAR) has been considered as a normal classification problem in computer vision and pattern recognition. Due to the vast array of potential uses for HAR as well as its increasing demand, researchers' attention has been sparked. The effectiveness of vision-based HAR techniques versus sensor-based HAR techniques is another topic of continuing discussion.

In this project, we have made use of the Human activity data collected using ambient sensors. The dataset includes 32 features and 17 target classes. Data has been collected continuously while residents perform their normal routines.
Ambient PIR motion sensors, door/temperature sensors, and light switch

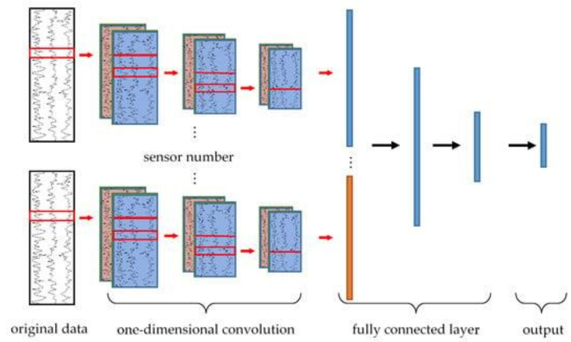
sensors
are placed throughout the home of the volunteer. The sensors are placed in locations
throughout the home that are related to specific activities of daily living that we wish to capture.

Background

The classification task is to predict the activity that is occurring in the smart home and being observed by the ambient sensors. The sensors communicate using the
ZigBee Pro protocol, forming a mesh network with all battery powered sensors as leaf
nodes and always-on devices. Artificial neural networks called convolutional neural networks were created to process structured matrices, including pictures. They are used for categorizing pictures based on the patterns and things that appear in them, including lines, circles, or even eyes and faces.

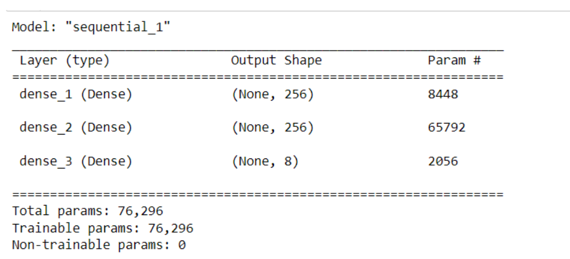
Approach

CNNs are being used extensively in a variety of tasks and activities, including speech recognition in the field of NLP, video analysis, and to resolve difficulties with image processing, computer vision, and self-driving car obstacle detection. Due to their major contribution to these rapidly developing and expanding fields, CNNs are widely used in deep learning.



We have first combined the data from each of the volunteers' dataset. We have then scaled and normalized the dataset to ensure the values are in the range of 0 to 1. We have then gone ahead to split the data into training and test, and validation datasets. For the purpose of activity classification, we have used eight of the seventeen classes available. We have then created a multilayer perceptron model with 4 layers and compared it with a 1D CNN model with 3 layers.

The model summary with number of layers and trainable parameters for the multilayer perceptron model are shown in the image below:



```
Model: "sequential_1"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 dense_1 (Dense)             (None, 256)               8448

 dense_2 (Dense)             (None, 256)               65792

 dense_3 (Dense)             (None, 8)                 2056

=================================================================
Total params: 76,296
Trainable params: 76,296
Non-trainable params: 0
_________________________________________________________________
```

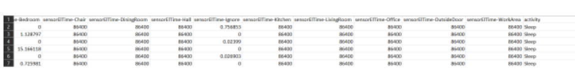
The model summary with number of layers and trainable parameters for the CNN model are shown in the image below:

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv1d (Conv1D)             (None, 30, 32)            128

 max_pooling1d (MaxPooling1D  (None, 15, 32)           0
 )

 flatten (Flatten)           (None, 480)               0

 dense (Dense)               (None, 8)                 3848

=================================================================
Total params: 3,976
Trainable params: 3,976
Non-trainable params: 0
```

Since the features are non-linear, we have used the Relu activation function to achieve non-linearity. We have also used a Maxpooling layer after the convolution layer to down sample the dataframe. We have also made use of the Adam optimizer to update the weights and learning rate accordingly. To evaluate the model performances, we have used "Accuracy" as our performance metric.

Results

The Human activity recognition dataset compiled using data collected from ambient continuous sensor data was used for the purpose of this classification task. The final dataframe consisted of 688743 rows of sensor data with 32 different features and "activity" being the target variable for this classification task.

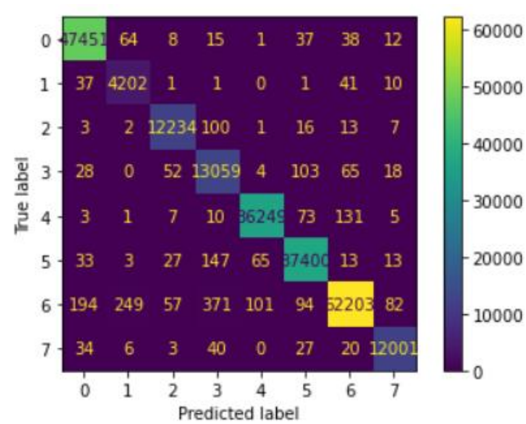


Furthermore, the training dataset consists of 67% of the data, while the test dataset contains the remaining 33%. We then normalized the split train and test datasets and used a multilayer perceptron model and a convolutional neural network to perform a classification task.

The accuracy of the model on the multilayer perceptron model for each epoch is shown in the image below:

```
In [118]:  1 MLP.fit(X_train, y_train, epochs=20, batch_size = 100)

Epoch 1/20
10329/10329 [==============================] - 17s 2ms/step - loss: 0.0109 - auc: 0.9998 - precision: 0.9966 - recall: 0.9965 -
Accuracy: 0.9965
Epoch 2/20
10329/10329 [==============================] - 16s 2ms/step - loss: 0.0103 - auc: 0.9998 - precision: 0.9967 - recall: 0.9966 -
Accuracy: 0.9967
Epoch 3/20
10329/10329 [==============================] - 16s 2ms/step - loss: 0.0102 - auc: 0.9999 - precision: 0.9968 - recall: 0.9967 -
Accuracy: 0.9967
Epoch 4/20
10329/10329 [==============================] - 16s 2ms/step - loss: 0.0101 - auc: 0.9998 - precision: 0.9968 - recall: 0.9968 -
Accuracy: 0.9968
Epoch 5/20
10329/10329 [==============================] - 16s 2ms/step - loss: 0.0100 - auc: 0.9998 - precision: 0.9969 - recall: 0.9968 -
Accuracy: 0.9968
Epoch 6/20
10329/10329 [==============================] - 16s 2ms/step - loss: 0.0101 - auc: 0.9998 - precision: 0.9970 - recall: 0.9969 -
Accuracy: 0.9969
Epoch 7/20
10329/10329 [==============================] - 16s 2ms/step - loss: 0.0096 - auc: 0.9998 - precision: 0.9970 - recall: 0.9970 -
Accuracy: 0.9970
Epoch 8/20
10329/10329 [==============================] - 16s 2ms/step - loss: 0.0094 - auc: 0.9999 - precision: 0.9971 - recall: 0.9970 -
Accuracy: 0.9970
Epoch 9/20
10329/10329 [==============================] - 16s 2ms/step - loss: 0.0093 - auc: 0.9998 - precision: 0.9971 - recall: 0.9971 -
Accuracy: 0.9971
Epoch 10/20
10329/10329 [==============================] - 17s 2ms/step - loss: 0.0095 - auc: 0.9998 - precision: 0.9971 - recall: 0.9971 -
Accuracy: 0.9971
Epoch 11/20
10329/10329 [==============================] - 16s 2ms/step - loss: 0.0090 - auc: 0.9998 - precision: 0.9972 - recall: 0.9972 -
Accuracy: 0.9972
Epoch 12/20
10329/10329 [==============================] - 16s 2ms/step - loss: 0.0093 - auc: 0.9998 - precision: 0.9972 - recall: 0.9971 -
Accuracy: 0.9971
Epoch 13/20
10329/10329 [==============================] - 16s 2ms/step - loss: 0.0091 - auc: 0.9998 - precision: 0.9972 - recall: 0.9972 -
Accuracy: 0.9972
Epoch 14/20
10329/10329 [==============================] - 16s 2ms/step - loss: 0.0088 - auc: 0.9998 - precision: 0.9973 - recall: 0.9973 -
Accuracy: 0.9973
Epoch 15/20
10329/10329 [==============================] - 16s 2ms/step - loss: 0.0089 - auc: 0.9998 - precision: 0.9973 - recall: 0.9973 -
Accuracy: 0.9973
Epoch 16/20
10329/10329 [==============================] - 16s 2ms/step - loss: 0.0085 - auc: 0.9998 - precision: 0.9974 - recall: 0.9974 -
Accuracy: 0.9974
Epoch 17/20
10329/10329 [==============================] - 16s 2ms/step - loss: 0.0085 - auc: 0.9998 - precision: 0.9974 - recall: 0.9973 -
Accuracy: 0.9973
Epoch 18/20
10329/10329 [==============================] - 16s 2ms/step - loss: 0.0088 - auc: 0.9998 - precision: 0.9974 - recall: 0.9973 -
Accuracy: 0.9974
Epoch 19/20
10329/10329 [==============================] - 18s 2ms/step - loss: 0.0087 - auc: 0.9999 - precision: 0.9974 - recall: 0.9974 -
Accuracy: 0.9974
Epoch 20/20
10329/10329 [==============================] - 17s 2ms/step - loss: 0.0084 - auc: 0.9998 - precision: 0.9975 - recall: 0.9975 -
Accuracy: 0.9975
Out[118]: <keras.callbacks.History at 0x1e5126d3760>
```

Below is an image of the confusion matrix generated:



Conclusions

The project results encouraged us to believe that neural networks are indeed performing well to classify human activity. The multilayer perceptron model seems to outperform the CNN for this particular classification task greatly in terms of accuracy with no visible underfitting or overfitting.

There are tremendous advantages to using HAR in a variety of application domains, but there are several unresolved issues and limitations with current HAR solutions that need to be resolved which are discussed below. Data-oriented research relies heavily on the data collection process. According to several existing research, there are a variety of problems with data collecting, including unlabeled datasets, a lack of temporal knowledge, unidentified class detection, and data restrictions that need to be resolved for activity recognition and prediction. ML and deep learning both heavily rely on data preprocessing. It is essential to preprocess data after data collection in order to extract useful information for HAR. "Appearance and feature extraction" and "background reduction" are two problems with data preprocessing, according to several research. Data annotation by hand takes time, produces erroneous labels, and muddies the chronology of events. The accuracy of HAR may be lowered by a poorly aligned dataset annotation. Additionally, recurring instability in prediction can be shown due to frame limit uncertainty if the action frame length is small. When the frame of one action is split into two or more distinct frames, misalignment happens, which causes the loss of some important information during frame segmentation. Additionally, misaligned activities result in inaccurate action detection, which lowers the effectiveness of HAR solutions. Getting past these underlying difficulties would comprise the future work segments.

Discussion and Future Work

A variety of situations, including real-time activities, group-based activities, and daily life activities, have been covered in the HAR literature and research that already exists. This research found that the bulk of previous studies concentrated on user behaviors as well as static and dynamic living activities. There are, however, surprisingly few studies on real-time activity recognition for a range of uses, including healthcare, surveillance, and suspicious conduct. The intricacy of real-time activities, hardware and technical limitations, and the scarcity of data are the main causes of the small number of research. Real-time activities require considerable computational power because background settings and context quickly change in busy or public spaces.

References

1. Köping, L.; Shirahama, K.; Grzegorzek, M. A general framework for sensor-based human activity recognition. Comput. Biol. Med. 2018, 95, 248–260.
2. Gupta, N.; Gupta, S.K.; Pathak, R.K.; Jain, V.; Rashidi, P.; Suri, J.S. Human activity recognition in artificial intelligence framework: A narrative review. Artif. Intell. Rev. 2022, 55, 4755–4808.
3. Khan, Z.N.; Ahmad, J. Attention induced multi-head convolutional neural network for human activity recognition. Appl. Soft Comput. 2021, 110, 107671.