Datathon 5:
Enhancing Mobile Health (mHealth) with Recurrent Neural Networks and Long Short-Term Memory (LSTM): A Deep Learning Approach to Activity Recognition and Predictive Modeling Using Wearable Sensor Data

Sashini Kosgoda, Ente Kang, Tamneet Tiwana (Team 3)

Dalla Lana School of Public Health, University of Toronto

CHL 5230: Applied Machine Learning for Health Data

Dr. Zahra Shakeri

November 21st, 2023

Introduction

The use of mobile health (mHealth) technologies in modern healthcare has been transformative, particularly in improving personalized healthcare management and disease prevention. Central to this transformation is the ability to accurately classify and predict human activities through wearable sensors. This report aims to address two primary questions: First, how accurately can sensor data classify different physical activities, such as walking, jogging, and climbing stairs? Second, can a LSTM model be developed that utilizes accelerometer and gyroscope readings to identify specific activities? Using the machine learning pipeline process, we will outline the data pre-processing and prediction techniques and evaluate the performance of the LSTM model in predicting activities using gyroscope and accelerometer readings. Addressing these questions is significant as it can lead to advancements in remote monitoring of patients, support for elderly care, and optimization of personal fitness programs.

Exploratory analysis and data processing

The mHealth dataset contained 999,999 observations with 14 variables. There were a total of 9 unique patients, each of which had a different number of repeated measurements. We started by checking to see if any of the variables had missing values, and confirmed that the dataset contained all complete cases. The activity variable contained a total of 12 unique activities that an individual may be classified as doing and 1 no activity class. We looked at the distribution of the activities among each individual, and saw that everyone spent most of their time doing nothing (activity = 0). This can pose problems with class imbalance later down the road, where oversampling may not be adequate to address. As such, we decided to do another analysis by dropping the observations with no activity since we are interested in building a model to predict activity of individuals. This new data frame consisted of 279,053 observations. We then looked at histograms to see the distribution of the features across the activities. Next, the correlation matrix was computed to see the correlation structure between different variables in our study including predictors and our response variable (activity).

Since there were repeated measurements for 9 unique individuals, the train-test split will have to be done by splitting individuals to train or test. This is because we do not want individuals' observations to be in both the train and test set, which will cause complications. Additionally, we normalized all the input features and created sequences to then be used for the LSTM.

Analysis

After preprocessing our data, two neural network models were created, with one model including activity=0, and one without activity=0. Neural network architecture featuring LSTM layers was employed due to its proven efficacy in handling sequential data. To prevent overfitting, dropout layers were introduced following the LSTM layers. The network's architecture was further comprised of fully connected layers that incorporated the Rectified Linear Unit activation functions and additional dropout layers to facilitate non-linear learning. Next, L2 regularization was implemented to mitigate overfitting in the model.

The Adam optimizer and the Cross-Entropy Loss function were used in training as it is suited for multi-class classification tasks. The hyperparameters set for the model including the activity=0, contained 13 output classes, a hidden layer size of 256, a batch size of 512, and a learning rate of 0.001, with a scheduling strategy implemented to adjust the learning rate throughout the training epochs. Similarly, the model without activity=0 had the same set of hyperparameters, however it only contained 12 output classes. The model was trained over 30 epochs to optimize the model's convergence speed and accuracy. The model's performance was evaluated on the training and validation set to assess its accuracy. Additionally, training and validation accuracies were plotted over epochs to visually interpret the model's learning progression, identify patterns, and generalizability when applied to unseen data.

Findings

Our exploratory analysis shows that the classes of activity are balanced where there was activity and not balanced in the class where individuals had no activity. Further, there were no missing values found in the dataset and After training the model without activity=0, we found that there was an average training accuracy of 60.11% and an average testing accuracy of 41.68%. This means the model without activity=0 correctly predicts the type of activity of an individual using sensor data 41.68% of the time, which shows a low performance. Further, when including the no activity class (activity = 0) the average training accuracy was 74.32% and the average testing accuracy was 74.60%. This means that the model with activity=0 correctly predicts the type of activity of an individual using sensor data 74.60% of the time, which indicates a moderate performance. The model without activity=0 showcased a generalization gap between the training and validation accuracy curve. With the training accuracy peaking at approximately 80% and the validation accuracy hovering around 40%, this gap indicates the model is not generalizing and is performing better on the training set than on unseen data. In addition to this model, the absence of convergence in the accuracy curves for both training and validation indicates that the model has not reached an optimal learning state. Overall, for both models, we also found that the accuracy of both training and testing sets is increasing as the number of epochs increases.

Limitations and Conclusion

In conclusion, our study underscores the nuanced considerations in developing RNN and LSTM models for activity recognition using wearable sensors. The observed class imbalance, particularly when including ‘no activity’ in the model may have skewed the performance metrics. While the model which included activity =0 did perform at a higher rate than guessing, their generalizability to unseen data is a critical factor in determining their practical utility. These models could serve as valuable tools for foundational research, highlighting the potential of RNNs in predicting activities with wearable sensors. These findings can also act as groundwork for prevention, management, and potential integration in patient monitoring/ fitness apps. This work accentuates the growing importance of personalized healthcare solutions in the realm of mHealth, where cutting-edge technologies hold transformative potential for enhancing individual well-being.

Team and individual contributions

All authors contributed equally to each section of this report, code, and presentation.

Code and Presentation

Github link: https://github.com/entekang/3-CHL5230-F23

Please find a copy of our raw code with notes at:
https://colab.research.google.com/drive/1ph5UyNDsDkhlSPeyHVH82Ya5nJHK4ZML?authuser=1#scrollTo=qZ1I8Woy_Jsb

Please find a copy of our google presentation at the following link:
https://docs.google.com/presentation/d/1G9Mv-qpJa0gEJk43irNxonkD0HsNg6IBOYApnpcpHQ8/edit?usp=sharing