Machine Learning

CLASSYFIRES

April 3, 2017

Abstract

Human Activity Recognition using Smartphone accelerometer and gyroscope data

1 Introduction

There are six tasks to classify: Stand, sit, walk, bike, stairs-up and stairs-down. The dataset contains the accelerometer and gyroscope readings in the x, y and z direction sampled at an interval of 100 Hz.The activities were carried out by 9 users on 4 smartphone models,namely Nexus4, Samsung S3, Samsung Gold and Samsung S3 mini.The classification was done using neural networks separately trained on the accelerometer data and the gyroscope data.

2 Procedure

2.1 Data processing

In general, data produced by sensors generate a large number of data. The dataset was partitioned into 13 partitions for readability purposes and was processed by taking out samples at the rate of 10Hz.The attribute 'Model' is given integer value from 1 to 4 and the attribute 'User' is given polar values. Since in the final model we need the user bias to be as minimal as possible, the values are set close to 0. The 13 compressed files are merged to form the final dataset.

2.2 Classification

Multi Layer Perceptron Classifier is used for classification. The library used is scikit-learn. The classification is done using neural networks, training separately on the accelerometer data and the gyroscope data. The number of hidden layers was varied from 1 to 3 and the units in each layer are varied in multiples of 5: 5, 10 and 15.The classification metric was imported from scikit-learn.

2.3 Training and Cross validation

The model is trained on the data obtained from 8 users. The 9th user is used for cross validating the model obtained.

activity	precision	recall	f1-score	samples
Stand	0.76	0.94	0.84	163997
Sit	0.95	0.97	0.96	175707
Walk	0.42	0.41	0.42	193213
Stairs Up	0.40	0.45	0.42	157562
Stairs Down	0.47	0.12	0.19	141931
Bike	0.49	0.62	0.55	161310
avg	0.58	0.60	0.57	993720

Table 1: Accelerometer Train Scores (3 Hidden layers and 5 hidden units)

activity	precision	recall	f1-score	samples
Stand	0.36	0.50	0.42	21152
Sit	1.00	1.00	1.00	23483
Walk	0.45	0.42	0.43	26023
Stairs Up	0.59	0.24	0.34	20641
Stairs Down	0.68	0.01	0.03	19653
Bike	0.25	0.52	0.34	23244
avg	0.55	0.46	0.44	134196

Table 2: Accelerometer Test Scores (3 Hidden layers and 5 hidden units)

3 Results and Interpretation

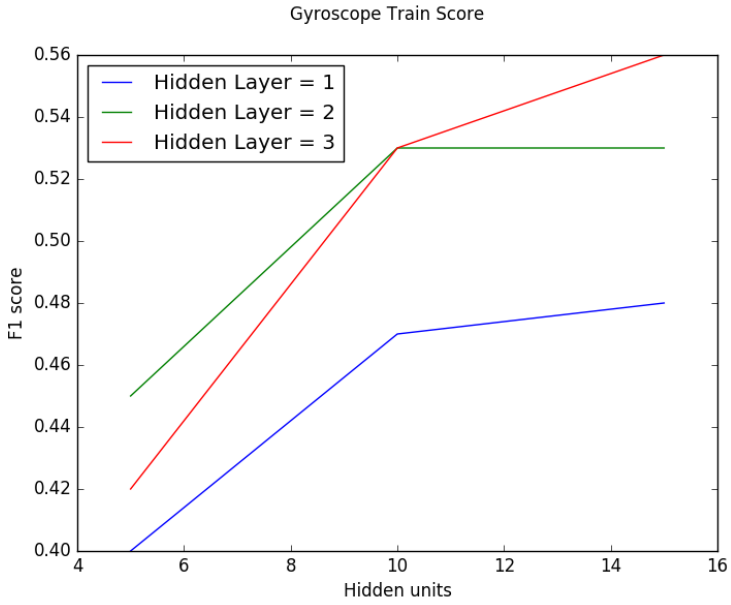
3.1 Model Accuracy

There are 9 neural network architecture for each of the Gyroscope and Accelerometer model. For Gyroscope model of Human Activity Recognition, the network architecture which gave the best result is 2 hidden layers with 10 units in each layer and that for Accelerometer data is 3 hidden layers with 5 units in each layer.

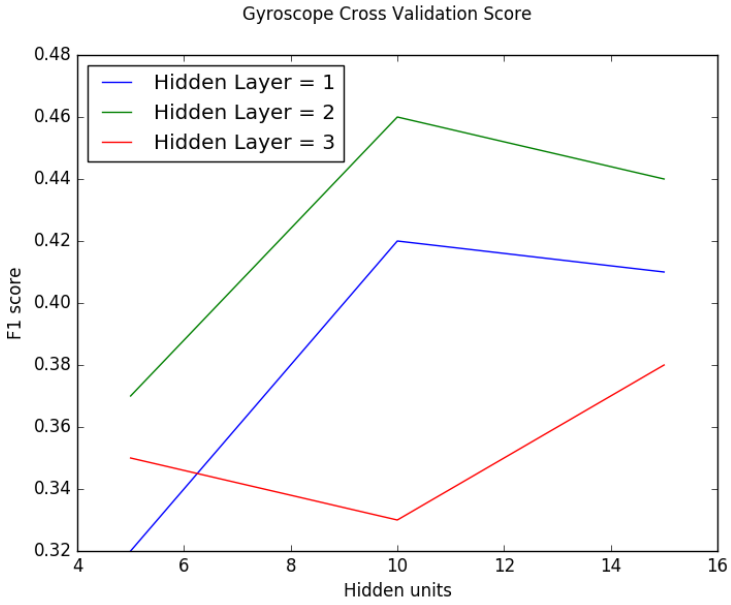
Detailed Classification metric for the above two models obtained from Gyroscope and Accelerometer is provided in the table.

activity	precision	recall	f1-score	samples
Stand	0.63	0.47	0.54	179074
Sit	0.64	0.82	0.71	195091
Walk	0.41	0.61	0.49	206735
Stairs Up	0.55	0.37	0.44	166708
Stairs Down	0.56	0.36	0.44	146885
Bike	0.51	0.49	0.50	166411
avg	0.55	0.53	0.53	1060904

Table 3: Gyroscope Train Scores (3 Hidden layers and 10 hidden units)

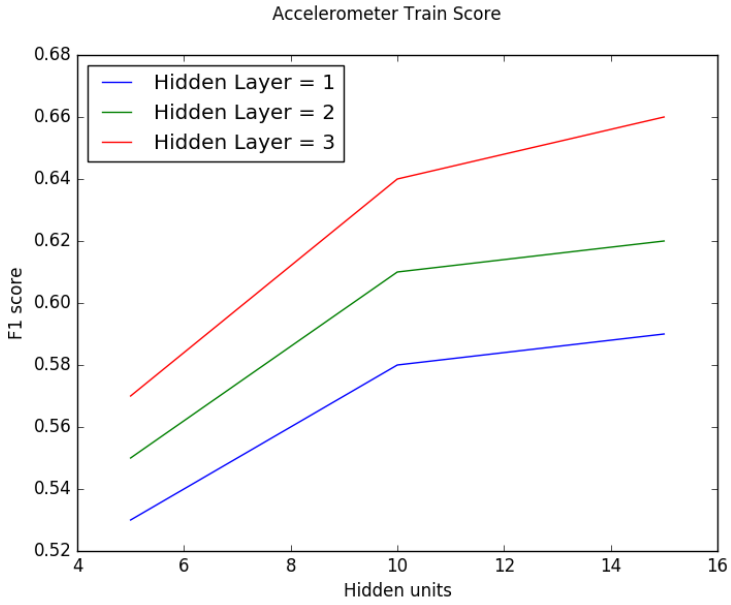


(a) Train Data.

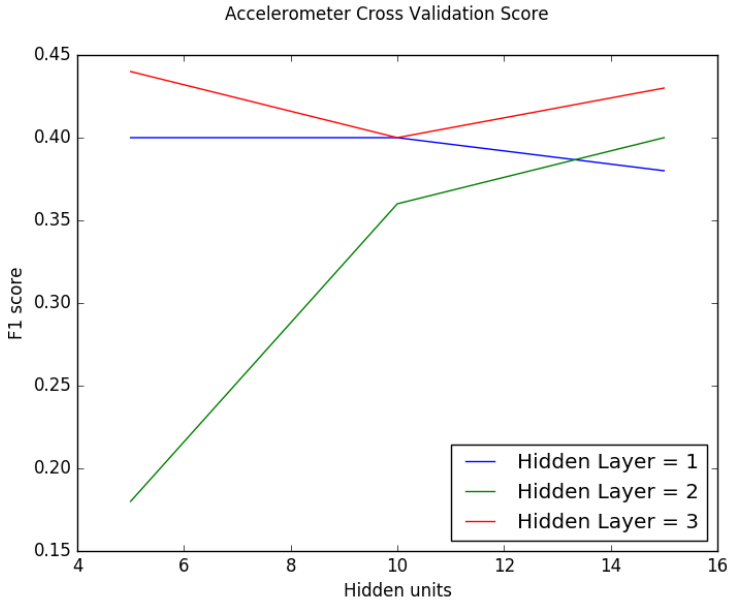


(b) CV Data.

Figure 1: Gyroscope F1 scores



(a) Train Data.



(b) CV Data.

Figure 2: Accelerometer F1 scores

activity	precision	recall	f1-score	samples
Stand	0.53	0.44	0.48	23345
Sit	0.64	0.88	0.74	26759
Walk	0.35	0.69	0.46	28308
Stairs Up	0.73	0.22	0.33	21730
Stairs Down	0.75	0.25	0.38	20491
Bike	0.31	0.25	0.28	24761
avg	0.54	0.48	0.46	145394

Table 4: Gyroscope Test Scores (3 Hidden layers and 10 hidden units)

3.2 Observations

The F1 score for the training data for Gyroscope and Accelerometer is 0.53 and 0.57 respectively. This implies that the best Neural Network model for both the sensor data is not able to fit the data through which it is trained. The F1 score for the test data for Gyroscope and Accelerometer is 0.46 and 0.44.

One main observation is that the model trained from Accelerometer data is unable to confidently classify the "Stairsdown" activity. As expected, the model trained from Accelerometer identifies the "Sit" activity with 100% accuracy. One of the main reason for misclassification of the "Stairsdown" activity using only accelerometer data is at any instance of time "Standing" and "Stairsdown" provide similar sensor readings.

There is no such bias in the misclassification of activities obtained from the model trained using Gyroscope data only, though "Sit" activity still has the highest F1 score.

4 Conclusion

Models trained from Gyroscope or Accelerometer alone are not able to fit themselves and the test data too. There has to be a clubbing of the two sensor datasets in order to obtain a higher rate of classification.

5 References

- Allan Stisen, Henrik Blunck, Sourav Bhattacharya, Thor Siiger Prentow, Mikkel Baun Kjærgaard, Anind Dey, Tobias Sonne, and Mads Møller Jensen "Smart Devices are Different: Assessing and Mitigating Mobile Sensing Heterogeneities for Activity Recognition" In Proc. 13th ACM Conference on Embedded Networked Sensor Systems (SenSys 2015), Seoul, Korea, 2015.

- Using Machine Learning on Sensor Data (Journal of Computing and Information Technology - CIT 18, 2010, 4, 341–347 doi:10.2498/cit.1001913)