Project Report
CSE-572 Data Mining
Spring -2018

Submitted to
Prof. Ayan Banerjee
IRA A. Fulton School of Engineering
Arizona State University

Submitted By

Venkata Ravi Teja Reddy Yanamala (1213420206)

Sai Nageswara Koushik Dhumantarao (1213158581)

Shreyas Shrinivas Mugali (1212904028)

Atul Mishra (1213363981)

Yeswanth Chowdary Narra (1213523673)

Table of Contents

Introduction:

This project is part of the Data Mining course taken by Dr. Ayan Banerjee in the Spring of 2018. The project solves the issue of a machine-level understanding of gesture recognition with the help of data from different sensors. This report is in continuation to previous assignments where we had found the PCA results of the data feature matrix. Here, we will do a user-dependent analysis on the new data feature matrix found by multiplying the PCA matrix with the new feature set matrix where 60% of each user data is used as training and rest 40% is used as test data. The training is done in three different machines: a) decision tree, b) support vector machine, and c) neural network. Finally, we will go on to report the accuracy metrics of precision, recall and F1 score for the test data.

5. Assignment 3

The third assignment broadly comprises of the following three tasks: New Feature Matrix, Training, and Testing. Following sections describe these tasks in more detail.

5.1. Task 1: New Feature Matrix

In this task we need to generate the new feature matrix. We first extract the 10 different gestures of each user in 10 different csv files. Then we apply the feature extraction methods on these csv files and multiply these matrix with the first five columns of the PCA matrix we generated earlier. The result is a set 10 csv files for each user. Now we append all the 10 csv files of each user and add an extra label column which identifies the different gestures. We have used the following labels-

1- About
2- And
3- Can
4- Cop
5- Deaf
6- Decide
7- Father
8- Find
9- Go Out
10- Hearing

This gives us 1 csv file per user. Finally, we shuffle these appended matrices which contains all the gesture features and the corresponding gesture labels. This is our new feature matrix which will now be used to train and test the machines.

5.2. Task 2: Training

Here, we need to train three different machines with user data. For each user we use a method called dividerand() to divide the new feature matrix into 60-40 partition. Now we use this 60% partition to train the machines in following manner.

5.2.1. Machine 1: Decision Tree

We used the fitctree() function from MATLAB library to train the decision tree.

5.2.2. Machine 2: Support Vector Machine

We used the fitcsvm() function from MATLAB library to train the support vector machine.

5.2.3. Machine 3: Neural Network

We used patternnet() function to define a neural network with 5 hidden layers and 10 neurons in each hidden layer. This neural network is then trained using train() function from the MATLAB library.

5.3. Task 3: Testing

In this task we test the machines by using rest 40% of the user data and generating the accuracy metrics describing precision, recall and F1 score of the trained machine. We calculated the performance metrics for all the users.

5.3.1. Precision

Precision is defined as- (true positives) / (true positives + false positives). When we tested our machines, we used user-defined MATLAB functions to calculate precision for all the users. Each user got a separate precision value for different gestures. We took the mean of all gesture precisions per user and reported this mean as the precision value for that user.

5.3.2. Recall

Recall is defined as- (true positives) / (true positives + false negatives). When we tested our machines, we used user-defined MATLAB functions to calculate recall for all the users. Each user got a separate recall value for different gestures. We took the mean of all gesture recall per user and reported this mean as the recall value for that user.

5.3.3. F1 Score

F1 score is defined as- (2 * precision * recall) / (precision + recall). When we tested our machines, we used user-defined MATLAB functions to calculate F1 score for all the users. Each user got a separate F1 score value for different gestures. We took the mean of all gesture F1 score per user and reported this mean as the F1 score value for that user.

5.3.4. Accuracy

Accuracy is defined as- (true positive + true negative) / (total samples). When we tested our machines, we used user-defined MATLAB functions to calculate F1 score for all the users. Each user got a separate accuracy value for different gestures. We took the mean of all gesture accuracy per user and reported this mean as the accuracy value for that user.

Following are the performance metrics for each user-

SVM:

User	Precision	Recall	F1 Score	Accuracy
User1	0.755754	0.745783	0.742338	0.950685
User2	0.738591	0.791802	0.746472	0.95283
User3	0	0	0	0
User4	0.77963	0.77963	0.77963	0.981481
User5	0.901169	0.891883	0.881474	0.975
User6	0.630966	0.633965	0.622661	0.935065
User7	0.605965	0.626313	0.609051	0.915
User8	0	0	0	0
User9	0.590012	0.581526	0.561602	0.914634
User10	0	0	0	0
User11	0.52881	0.537937	0.483984	0.905
User12	0.9	0.883333	0.88535	0.972727
User13	0.728135	0.70381	0.709034	0.9425
User14	0	0	0	0
User15	0.86	0.869697	0.852455	0.968293
User16	0.764394	0.741894	0.743658	0.947826
User17	0	0	0	0
User18	0	0	0	0
User19	0.631746	0.685335	0.630709	0.929268
User20	0.852698	0.861865	0.852174	0.973494
User21	0	0	0	0
User22	0.69869	0.700128	0.688702	0.936585
User23	0.870437	0.866944	0.861422	0.969863
User24	0.77197	0.792778	0.773175	0.955294
User25	0	0	0	0
User26	0.561299	0.532554	0.505718	0.907937
User27	0.695628	0.70267	0.691397	0.938095
User28	0.901934	0.898864	0.891873	0.975904
User29	0.771334	0.778889	0.750046	0.958025
User30	0	0	0	0
User31	0.8377	0.799475	0.80052	0.955844
User32	0.75118	0.739643	0.724491	0.951899
User33	0	0	0	0
User34	0.771843	0.764643	0.762958	0.9475
User35	0	0	0	0
User36	0.790159	0.784167	0.773987	0.95122
User37	0	0	0	0

To standardize all the data values, we have zero-padded each row to a fixed value of 55 columns in the previous assignment. If any row in an action file has more than 55 columns of time series then we have ignored that file. This resulted in some users not having all the 10 gesture csv files. Users with less than 10 gesture csv files have not been considered and thus reported to have 0 value for precision, recall, F1 score and accuracy.

Decision Tree:

User	Precision	Recall	F1 Score	Accuracy
User1	0.719722	0.715386	0.711317	0.939726
User2	0.693878	0.605249	0.568697	0.933962
User3	0	0	0	0
User4	0.699786	0.711905	0.695238	0.959259
User5	0.82776	0.800685	0.794601	0.955
User6	0.735462	0.715303	0.711541	0.94026
User7	0.564643	0.606288	0.564198	0.91
User8	0	0	0	0
User9	0.680281	0.687424	0.645201	0.929268
User10	0	0	0	0
User11	0.629946	0.622063	0.611355	0.9225
User12	0.869444	0.797222	0.785701	0.972727
User13	0.638712	0.600476	0.594922	0.9225
User14	0	0	0	0
User15	0.882078	0.9	0.876615	0.97561
User16	0.614808	0.558434	0.556247	0.913043
User17	0	0	0	0
User18	0	0	0	0
User19	0.542012	0.609812	0.53659	0.909756
User20	0.740797	0.723593	0.709876	0.942169
User21	0	0	0	0
User22	0.580198	0.584744	0.566167	0.912195
User23	0.802172	0.787024	0.776808	0.953425
User24	0.833824	0.781111	0.735771	0.952941
User25	0	0	0	0
User26	0.738333	0.731515	0.656202	0.936508
User27	0.621944	0.626407	0.614897	0.92619
User28	0.808214	0.792904	0.768048	0.954217
User29	0.45812	0.491429	0.446748	0.893827
User30	0	0	0	0
User31	0.71004	0.693144	0.676618	0.92987
User32	0.696665	0.638254	0.65916	0.939241
User33	0	0	0	0

User34	0.830942	0.819643	0.79794	0.9575
User35	0	0	0	0
User36	0.750325	0.68	0.67877	0.931707
User37	0	0	0	0

Users with less than 10 gesture csv files have not been considered and thus reported to have 0 value for precision, recall, F1 score and accuracy.

Neural Network:

User	Precision	Recall	F1 Score	Accuracy
User1	0.607468	0.606984	0.603038	0.915068
User2	0.355556	0.270833	0.157609	0.811321
User3	0	0	0	0
User4	0.554545	0.564755	0.503043	0.883333
User5	0.781667	0.805152	0.781663	0.9525
User6	0.562309	0.498763	0.479493	0.927273
User7	0.382228	0.39246	0.34112	0.87
User8	0	0	0	0
User9	0.421429	0.43184	0.353911	0.870732
User10	0	0	0	0
User11	0.611898	0.577619	0.526884	0.915
User12	0.842916	0.735238	0.727501	0.960606
User13	0.578413	0.54619	0.485052	0.9
User14	0	0	0	0
User15	0.732706	0.723492	0.691869	0.941463
User16	0.53228	0.538016	0.460364	0.902174
User17	0	0	0	0
User18	0	0	0	0
User19	0.381621	0.439881	0.320039	0.882927
User20	0.598748	0.637659	0.598141	0.922892
User21	0	0	0	0
User22	0.508	0.500804	0.465648	0.897561
User23	0.695379	0.669246	0.620779	0.942466
User24	0.688269	0.653214	0.648017	0.936471
User25	0	0	0	0
User26	0.519048	0.476111	0.448899	0.892063
User27	0.723739	0.713636	0.650705	0.930952
User28	0.780153	0.7463	0.745193	0.949398
User29	0.685211	0.671212	0.666471	0.938272
User30	0	0	0	0
User31	0.544938	0.556115	0.491473	0.901299

User32	0.318452	0.349513	0.26907	0.865823
User33	0	0	0	0
User34	0.593074	0.591865	0.532938	0.9175
User35	0	0	0	0
User36	0.654521	0.603644	0.572094	0.921951
User37	0	0	0	0

Users with less than 10 gesture csv files have not been considered and thus reported to have 0 value for precision, recall, F1 score and accuracy.

5.3.5. Results

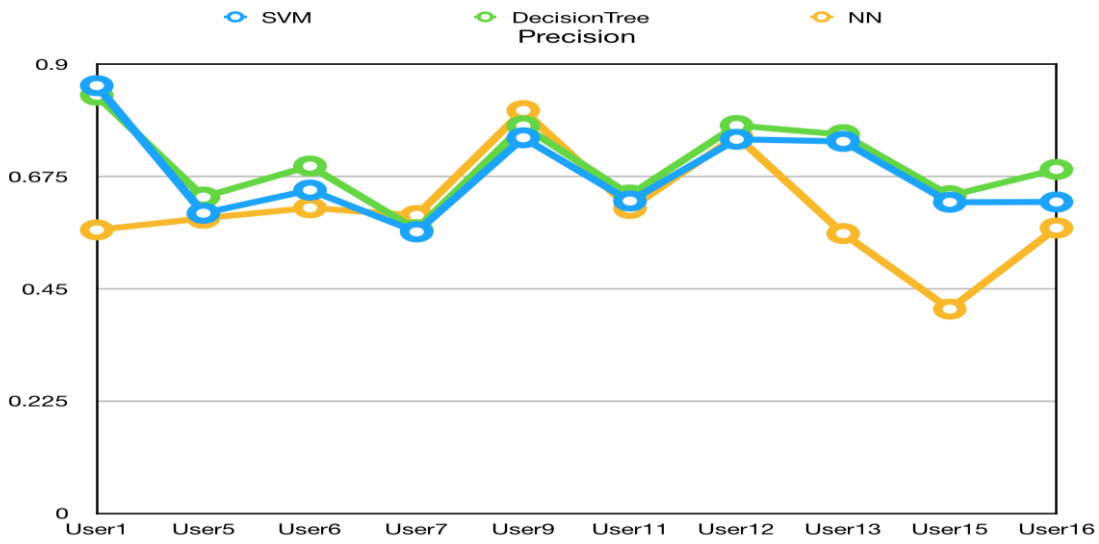

Fig: Plot showing precision for 10 random users in all machines

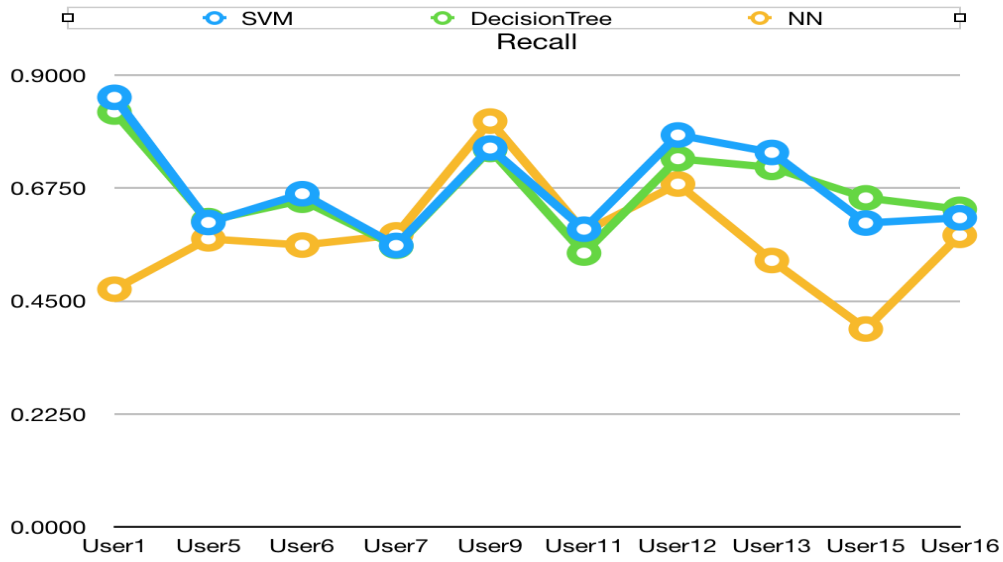

Fig: Plot showing recall for 10 random users in all machines

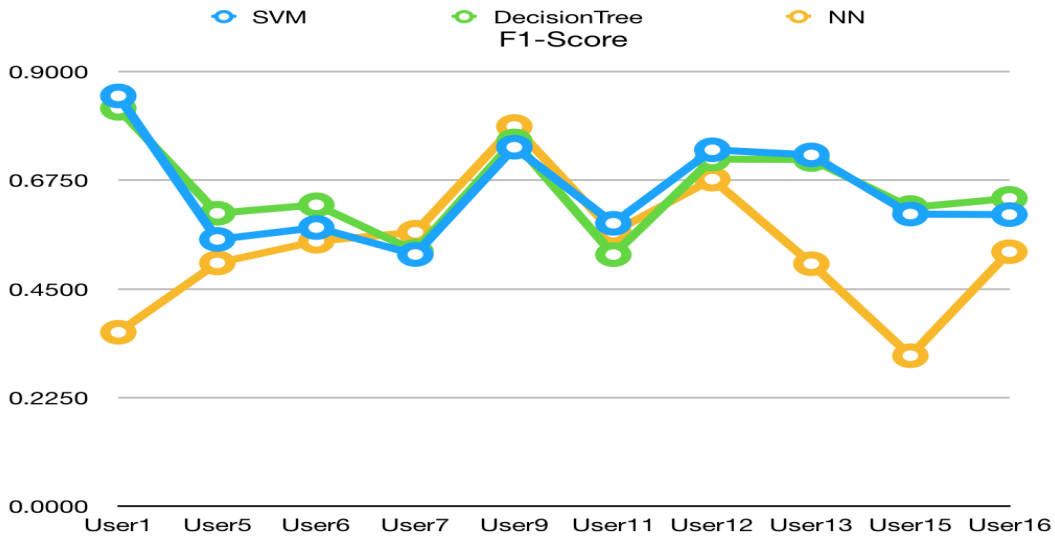

Fig: Plot showing F1 Score for 10 random users in all machines

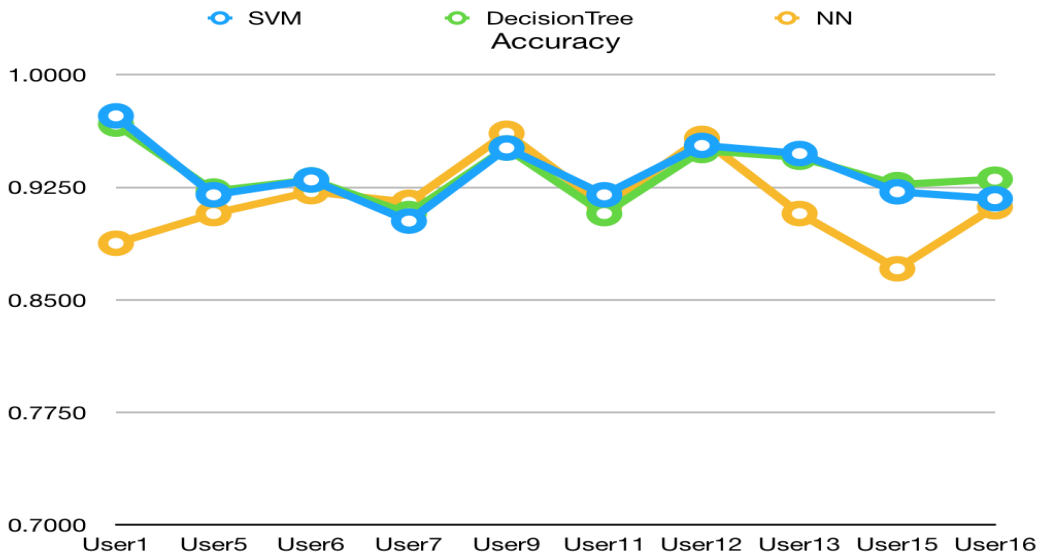

Fig: Plot showing accuracy for 10 random users in all machines

5.3.6. Conclusion

Since the given user data is less, we got better results in decision tree and svm machines as compared to neural network machine. This is reflected in the result plots shown above.