Group 24

Project 3 Report

Data Mining

CSE 572

Spring 2018

Submitted to:

Professor Ayan Banerjee

Ira A. Fulton School of Engineering

Arizona State University

Submitted by:

Anuhya Sai Nudurupati (anuduru@asu.edu)
Anvitha Dinesh Rao (arao30@asu.edu)
Rachana Kashyap (rnkashya@asu.edu**)**
Rohit Balachandran Menon (skolli6@asu.edu**)**
Vaishak Ramesh Vellore (vvellore@asu.edu**)**

1 Introduction

In this project, we attempt to develop a system which can understand and recognize the American Sign Language(ASL) through human gestures. A wristband sensor worn on both hands is used to collect data related to acceleration, gyroscope, orientation, electromyography and kinect data and is mined to understand what gesture the person has made. This could help a person who does not understand ASL to be able to communicate with a deaf/dumb person who does communicate in ASL. We use MATLAB to develop this software.

2 Project Phase 1

In the first phase, we went to the IMPACT lab at Brickyard, Tempe in order to collect data. One person wore wrist bands on both arms and made the gestures, “ABOUT”, “AND”, “CAN”, “COP” ,“DEAF”,“DECIDE”, “FATHER”, “FIND”, “GO OUT” and “HEARING” about 20 times each. The data collected from the sensors is stored in the form of CSV files. The time series data is sampled every 3 seconds. The frequency of sensors was found to be 15Hz. The data headers of the collected data are Accelerometer, Electromyogram, Gyroscope and Orientation.

3 Project Phase 2

The second phase of the project involves feature extraction and feature selection aspects of Data Mining. PCA was applied to the feature matrix to obtain the new feature matrix . From the feature matrix 7 features were extracted and multiplied with the feature matrix obtain a projection matrix . This projection matrix is used as a new feature matrix.

4 Project Phase 3

The third phase of the project involves the following steps,

A. A new column is added to the new feature matrix obtained from phase 2 for each user in order to create labels used for binary classification.
B. The data is shuffled and selected at random from the new feature matrix with labels generated.
C. 60% of the data for each user was used for training.
D. 40% of the remaining data was used for testing.
E. Support Vector Machines, Neural Networks and Decision Trees were used for training the machine .
F. The test dataset is then used to obtain the accuracy metrics Precision, Recall and F1 score for each user.

Code:

```
files = dir('*.csv');
i=1;
decision_tree_accuracy = zeros(1,10);
svm_res_accuracy = zeros(1,10);

for file=files'
csv=readtable(file.name,"ReadRowNames",false);

%%%%%%%%%%%%%%% Pre-processing %%%%%%%%%%%%%%%

random_dataset = csv(randperm(size(csv, 1)), :);
split = floor(size(random_dataset,1)/3);
var=split*2+1;
train_data = random_dataset(1:split*2,:);
test_data = random_dataset(var:end,1:end);
test_set = test_data(:,1:end-1);
test_labels = test_data(:,end:end);
train_set = train_data(:,1:end-1);
train_labels = train_data(:,end:end);
```

```
%%%%%%%%%%%% DECISION TREE %%%%%%%%%%%%%%%%%%
dt = fitctree(train_set,train_labels);
p = predict(dt, test_set);
decision_tree_accuracy(i) =
sum(table2array(test_labels)==p)/size(table2array(test_labels),1
;

[confMat1,order1] = confusionmat(table2array(test_labels), p);
recall_dt(i)=confMat1(1,1)/sum(confMat1(1,:));
precision_dt(i) = confMat1(1,1)/sum(confMat1(:,1));
f1_score_dt(i) =
2*recall_dt(i)*precision_dt(i)/(recall_dt(i)+precision_dt(i));

%%%%%%%%%%%%%%%%%% SVM %%%%%%%%%%%%%%%%%%%%%%%

svm = fitcecoc(train_set,train_labels);
v = predict(svm, test_set);
svm_res_accuracy(i) =
sum(table2array(test_labels)==v)/size(test_labels,1);

[confMat2,order2] = confusionmat(table2array(test_labels), v);
recall_svm(i)=confMat2(1,1)/sum(confMat2(1,:));
precision_svm(i) = confMat2(1,1)/sum(confMat2(:,1));
f1_score_svm(i) =
2*recall_svm(i)*precision_svm(i)/(recall_svm(i)+precision_svm(i)
;

%%%%%%%%%%%%%%%%%%% NEURAL NETS %%%%%%%%%%%%
net = patternnet(5);
net = train(net,table2array(train_set)',
table2array(train_labels)');
y = round(net(table2array(test_set)'));
[confMat3,order3] = confusionmat(table2array(test_labels), y');
recall_nn(i)=confMat3(1,1)/sum(confMat3(1,:));
precision_nn(i) = confMat3(1,1)/sum(confMat3(:,1));
f1_score_nn(i) =
2*recall_nn(i)*precision_nn(i)/(recall_nn(i)+precision_nn(i));
```

```
%%%%%%%%%%%%%%NEXT FILE ITERATION%%%%%%%%%%%%%%%
i= i+1;
end

%%%%%%%%%%%%AVERAGE R and P DT %%%%%%%%%%%%%%%%%
Average_Recall_User_DT = mean(recall_dt);
Average_Precision_User_DT = mean(precision_dt);

%%%%%%%%% AVERAGE R and P SVM %%%%%%%%%%%%%%%%%%

Average_Recall_User_SVM = mean(recall_svm);
Average_Precision_User_SVM = mean(precision_svm);

%%%%%%%%%%%%%%AVERAGE R and P NN %%%%%%%%%%%%%%%%
Average_Recall_User_NN = mean(recall_nn);
Average_Precision_User_NN = mean(precision_nn);

%%%%%%%%%%%% AVERAGE F1 DT and SVM and NN %%%%%%%
Average_F1_User_DT = mean(f1_score_dt);
Average_F1_User_SVM = mean(f1_score_svm);
Average_F1_User_NN = mean(f1_score_nn);
```

Accuracy Metrics for 10 Users:

Group 11:

Gesture	F1 DT	FT NN	F1 SVM	PRECISION DT	PRECISION NN	PRECISION SVM	RECALL DT	RECALL NN	RECALL SVM
About	0.9593496	0.992	0.9593496	0.967213115	0.984126984	0.967213115	0.9516129	1	0.9516129
And	0.975522	0.9731063	0.5529138	0.983949666	0.948306595	0.992307692	0.96774194	1	0.5245658
Can	1	1	0.8860878	1	1	0.944444444	1	1	0.8494624
Cop	0.9462133	0.936313	0.8323574	0.939946322	0.904799989	0.87934356	0.95422362	0.9712297	0.8162361
Deaf	0.961299	0.9565261	0.5206584	0.958536079	0.919852941	0.883755784	0.96480191	0.9967213	0.4137124
Decide	0.9759116	0.9535009	0.8030693	0.968291238	0.915958105	0.928858179	0.98391224	0.9945796	0.8167289
Father	0.9664317	0.9463049	0.7690761	0.971392768	0.908742724	0.882216231	0.96246921	0.9884025	0.7061227
Find	0.9544617	0.9379619	0.8194546	0.945468759	0.88915804	0.917769372	0.96496053	0.9939186	0.8316448
Goout	0.9611498	0.9502395	0.7939949	0.946524566	0.928978531	0.943148321	0.97688121	0.9735525	0.7213455
Hearing	0.9607664	0.9446695	0.8487328	0.955114336	0.90668307	0.890062871	0.96733383	0.9875412	0.8392983
Average	**0.9661105**	**0.9590622**	**0.7785695**	**0.963643685**	**0.930660698**	**0.922911957**	**0.96939374**	**0.9905945**	**0.747073**

Group 13:

Gesture	F1 DT	FT NN	F1 SVM	PRECISION DT	PRECISION NN	PRECISION SVM	RECALL DT	RECALL NN	RECALL SVM
About	0.9747899	0.974359	0.2739726	0.966666667	0.982758621	0.714285714	0.98305085	0.9661017	0.1694915
And	0.9789855	0.9477419	0.9379909	1	0.915584416	0.938461538	0.95912125	0.983871	0.9429922
Can	0.9779558	0.9438659	0.8862419	0.956895515	0.898445998	0.890106952	1	0.9943503	0.8856419
Cop	0.9811165	0.9257212	0.8128649	0.971038251	0.915303782	0.932539683	0.99180328	0.942257	0.7555035
Deaf	0.9718279	0.945917	0.856822	0.961315577	0.918894381	0.888931221	0.98315419	0.9762654	0.8403076
Decide	0.9944179	0.9956905	0.9261505	0.988972315	0.997354497	0.981481481	1	0.994152	0.8815043
Father	0.9843368	0.9434448	0.9322133	0.978525845	0.910031375	0.946863734	0.99078341	0.9814535	0.9226679
Find	0.9820464	0.959422	0.8995819	0.981174688	0.951625935	0.929797404	0.98353153	0.9712936	0.8809206
Goout	0.9851572	0.9662529	0.912955	0.983833891	0.95303214	0.961103853	0.98708094	0.9814636	0.8871993
Hearing	0.9782074	0.9670071	0.8045217	0.960813313	0.947716267	0.874118666	0.9966092	0.9881428	0.7788082
Average	**0.9808841**	**0.9569422**	**0.8243315**	**0.974923606**	**0.939074741**	**0.905769025**	**0.98751346**	**0.9779351**	**0.7945037**

Group 15:

Gesture	F1 DT	FT NN	F1 SVM	PRECISION DT	PRECISION NN	PRECISION SVM	RECALL DT	RECALL NN	RECALL SVM
About	1	0.992	0.2253521	1	1	1	1	0.984127	0.1269841
And	0.9434146	0.9375	0.9331857	0.935867896	0.902306649	0.901377727	0.95108408	0.9756742	0.9673453
Can	0.991734	0.9515099	0.9232345	0.994535519	0.931667718	0.921498599	0.9890681	0.9725806	0.9290323
Cop	0.9648219	0.9389318	0.9327657	0.955244349	0.895551935	0.916697859	0.97509484	0.987835	0.9499295
Deaf	0.9790157	0.9724599	0.8538901	0.978105632	0.969858512	0.946551704	0.98042499	0.9769231	0.7868946
Decide	0.9758402	0.962127	0.7475958	0.976052937	0.939363913	0.963494318	0.97588326	0.9866445	0.705867
Father	0.9780044	0.9495473	0.8751062	0.977102322	0.912637972	0.928454927	0.9793652	0.9907029	0.8401429
Find	0.9560713	0.9466558	0.9038113	0.950112133	0.898897059	0.89451476	0.9632958	1	0.9195512
Goout	0.9881479	0.8989351	0.8342981	0.988375066	0.982193038	0.985837983	0.98816211	0.9030976	0.7413425
Hearing	0.9662724	0.9682619	0.8257213	0.965589395	0.952303914	0.943196908	0.96798896	0.9853405	0.8013073
Average	**0.9743322**	**0.9517929**	**0.8054961**	**0.972098525**	**0.938478071**	**0.940162479**	**0.97703673**	**0.9762925**	**0.7768397**

Group 7:

Gesture	F1 DT	F1 NN	F1 SVM	PRECISION DT	PRECISION NN	PRECISION SVM	RECALL DT	RECALL NN	RECALL SVM
About	0.91071	0.910569	0.8440367	0.910714286	0.835820896	0.867924528	0.910714	1	0.82142857
And	0.97479	0.974359	0.9830508	0.966666667	0.982758621	0.983050847	0.983051	0.9661017	0.98305085
Can	0.9916	0.991597	0.928	1	1	0.892307692	0.983333	0.9833333	0.96666667
Cop	0.98387	0.983333	0.9365079	0.968253968	1	0.907692308	1	0.9672131	0.96721311
Deaf	0.99174	0.97479	0.9756098	0.983606557	0.983050847	0.952380952	1	0.9666667	1
Decide	0.98305	0.936508	0.974359	0.983050847	0.880597015	0.982758621	0.983051	1	0.96610169
Father	0.92562	0.919355	0.892562	0.875	0.850746269	0.84375	0.982456	1	0.94736842
Find	0.96774	0.944882	0.928	0.9375	0.895522388	0.892307692	1	1	0.96666667
GoOut	1	1	0.9345794	1	1	1	1	1	0.87719298
Hearing	0.98438	0.976	0.9767442	0.984375	1	0.969230769	0.984375	0.953125	0.984375
AVERAGE	**0.97135**	**0.961139**	**0.937345**	**0.960916733**	**0.942849604**	**0.929140341**	**0.982698**	**0.983644**	**0.9480064**

Group 16:

Gesture	F1 DT	F1 NN	F1 SVM	PRECISION DT	PRECISION NN	PRECISION SVM	RECALL DT	RECALL NN	RECALL SVM
About	0.94309	0.944	0.8214286	0.950819672	0.936507937	0.92	0.935484	0.9516129	0.74193548
And	1	1	0.9206349	1	1	0.935483871	1	1	0.90625
Can	0.9916	0.982759	0.8852459	0.983333333	1	0.857142857	1	0.9661017	0.91525424
Cop	0.98305	0.966102	0.9464286	0.983050847	0.966101695	1	0.983051	0.9661017	0.89830508
Deaf	0.992	0.96875	0.9411765	0.984126984	0.939393939	0.98245614	1	1	0.90322581
Decide	0.98413	1	0.9375	0.96875	1	0.909090909	1	1	0.96774194
Father	0.89076	0.929134	0.66	0.883333333	0.867647059	0.804878049	0.898305	1	0.55932203
Find	0.98333	0.991597	0.9655172	0.967213115	0.983333333	0.98245614	1	1	0.94915254
GoOut	0.976	1	0.9593496	0.968253968	1	0.967213115	0.983871	1	0.9516129
Hearing	0.97561	1	0.9586777	0.967741935	1	0.966666667	0.983607	1	0.95081967
AVERAGE	**0.97196**	**0.978234**	**0.8995959**	**0.965662319**	**0.969298396**	**0.932538775**	**0.978432**	**0.9883816**	**0.87436197**

Group 19:

Gesture	F1 DT	F1 NN	F1 SVM	PRECISION DT	PRECISION NN	PRECISION SVM	RECALL DT	RECALL NN	RECALL SVM
About	1	0.990991	0.990991	1	0.982142857	0.982142857	1	1	1
And	0.98276	0.991453	0.9491525	1	1	0.949152542	0.966102	0.9830508	0.94915254
Can	0.95495	0.99115	0.9454545	0.981481481	1	0.981132075	0.929825	0.9824561	0.9122807
Cop	0.97248	0.878505	0.2769231	1	0.921568627	1	0.946429	0.8392857	0.16071429
Deaf	0.99065	0.931034	0.9122807	1	0.870967742	0.866666667	0.981481	1	0.96296296
Decide	0.98246	1	0.6428571	0.98245614	1	1	0.982456	1	0.47368421
Father	0.98148	0.971963	0.9217391	0.981481481	0.981132075	0.868852459	0.981481	0.962963	0.98148148
Find	0.9913	0.991453	0.7173913	1	0.983050847	0.970588235	0.982759	1	0.56896552
GoOut	1	0.982143	0.9369369	1	1	0.962962963	1	0.9649123	0.9122807
Hearing	1	0.964286	0.9357798	1	0.981818182	0.980769231	1	0.9473684	0.89473684
AVERAGE	**0.98561**	**0.969298**	**0.8229506**	**0.99454191**	**0.972068033**	**0.956226703**	**0.977053**	**0.9680036**	**0.78162592**

Group 28:

Gesture	F1 DT	F1 NN	F1 SVM	Precision DT	Precision NN	Precision SVM	Recall DT	Recall NN	Recall SVM
About	0.93913	0.912	0.816327	0.93103448	0.838235294	0.975609756	0.947368	1	0.7017544
And	0.949153	0.9375	0.885246	0.96551724	0.882352941	0.870967742	0.933333	1	0.9
Can	0.96	0.944882	0.197183	0.96774194	0.9375	0.875	0.952381	0.952381	0.1111111
Cop	0.967213	0.967213	0.967742	0.9516129	0.951612903	0.9375	0.983333	0.983333	1
Deaf	0.992248	0.984375	0.977099	0.98461538	0.984375	0.955223881	1	0.984375	1
Decide	0.983871	0.96063	0.912	0.96825397	0.924242424	0.890625	1	1	0.9344262
Father	0.957265	0.920635	0.912	0.94915254	0.852941176	0.850746269	0.965517	1	0.9827586
Find	1	0.992126	0.913793	1	0.984375	1	1	1	0.8412698
Goout	0.983333	0.9375	0.929134	0.98333333	0.882352941	0.880597015	0.983333	1	0.9833333
Hearing	1	1	0.894309	1	1	0.873015873	1	1	0.9166667
AVERAGE	0.973221	0.955686	0.840483	0.97012618	0.923798768	0.910928554	0.976527	0.992009	0.837132

Group 34:

Gesture	F1 DT	F1 NN	F1 SVM	Precision DT	Precision NN	Precision SVM	Recall DT	Recall NN	Recall SVM
About	1	1	0.97561	1	1	0.983606557	1	1	0.9677419
And	0.966102	0.943089	0.940171	0.95	0.892307692	0.93220339	0.982759	1	0.9482759
Can	0.983871	0.945736	0.09375	0.96825397	0.897058824	1	1	1	0.0491803
Cop	1	0.983607	0.839286	1	1	0.94	1	0.967742	0.7580645
Deaf	0.977099	0.96063	0.117647	0.95522388	0.968253968	1	1	0.953125	0.0625
Decide	0.983333	0.99187	0.653465	1	0.983870968	0.825	0.967213	1	0.5409836
Father	0.982759	0.920635	0.862745	0.98275862	0.852941176	1	0.982759	1	0.7586207
Find	0.984375	0.952381	0.916667	0.96923077	0.952380952	0.964912281	1	0.952381	0.8730159
Goout	0.984127	0.983871	0.945736	0.98412698	1	0.924242424	0.984127	0.968254	0.968254
Hearing	1	0.983871	0.953846	1	1	0.925373134	1	0.968254	0.984127
AVERAGE	0.986167	0.966569	0.729892	0.98095942	0.954681358	0.949533779	0.991686	0.980976	0.6910764

Group 36:

Gesture	F1 DT	F1 NN	F1 SVM	Precision DT	Precision NN	Precision SVM	Recall DT	Recall NN	Recall SVM
About	0.991736	0.920354	0.151515	1	1	1	0.983607	0.852459	0.0819672
And	0.983607	0.99187	0.32	0.98360656	0.983870968	0.857142857	0.983607	1	0.1967213
Can	1	1	0.96	1	1	0.983606557	1	1	0.9375
Cop	0.967742	0.945736	0.92562	0.95238095	0.897058824	0.933333333	0.983607	1	0.9180328
Deaf	0.983607	0.9375	0.5	0.96774194	0.882352941	0.875	1	1	0.35
Decide	0.992126	0.953846	0.645833	0.984375	0.925373134	0.939393939	1	0.984127	0.4920635
Father	0.992	0.992	0.941176	0.98412698	0.984126984	0.98245614	1	1	0.9032258
Find	0.934426	0.9375	0.920635	0.91935484	0.882352941	0.878787879	0.95	1	0.9666667
Goout	0.890756	0.912	0.659794	0.88333333	0.863636364	0.842105263	0.898305	0.966102	0.5423729
Hearing	0.95	0.945736	0.920635	0.96610169	0.897058824	0.892307692	0.934426	1	0.9508197
AVERAGE	0.9686	0.953654	0.694521	0.96410213	0.931583098	0.918413366	0.973355	0.980269	0.633937

Group 29:

Gesture	F1 DT	FT NN	F1 SVM	PRECISION DT	PRECISION NN	PRECISION SVM	RECALL DT	RECALL NN	RECALL SVM
About	0.9827586	0.9913043	0.8	0.982758621	1	0.893617021	0.98275862	0.9827586	0.72413793
And	0.928	0.9302326	0.877193	0.892307692	0.869565217	0.925925926	0.96666667	1	0.83333333
Can	0.9430894	0.9365079	0.9302326	0.935483871	0.907692308	0.882352941	0.95081967	0.9672131	0.98360656
Cop	0.992	0.9465649	0.9302326	0.984126984	0.898550725	0.895522388	1	1	0.96774194
Deaf	0.983871	0.9606299	0.9612403	0.983870968	0.938461538	0.925373134	0.98387097	0.983871	1
Decide	0.9918699	0.976	0.7184466	0.983870968	0.953125	0.880952381	1	1	0.60655738
Father	0.968254	0.9365079	0.9152542	0.953125	0.921875	0.964285714	0.98387097	0.9516129	0.87096774
Find	0.8849558	0.9302326	0.1538462	0.943396226	0.869565217	1	0.83333333	1	0.08333333
Goout	0.9915966	0.9833333	0.9401709	1	0.983333333	0.964912281	0.98333333	0.9833333	0.91666667
Hearing	0.9917355	0.944	0.1538462	0.983606557	0.907692308	1	1	0.9833333	0.08333333
Average	**0.9658131**	**0.9535313**	**0.7380462**	**0.964254689**	**0.924986065**	**0.933294179**	**0.96846536**	**0.9852122**	**0.70696782**

Comparison Table of the average accuracy metrics for Decision Tree, Neural Networks and Support Vector Machine for 10 Users

	F1 DT	F1 NN	F1 SVM	PRECISION	PRECISION NN	PRECISION SVM	RECALL DT	RECALL NN	RECALL SVM
Group 7	0.971349	0.96114	0.937345	0.9609167	0.942849604	0.929140341	0.98269805	0.983644	0.9480064
Group 16	0.971956	0.97823	0.899596	0.9656623	0.969298396	0.932538775	0.97843173	0.9883816	0.87436197
Group 19	0.985609	0.9693	0.822951	0.9945419	0.972068033	0.956226703	0.97705326	0.9680036	0.78162592
Group 11	0.966111	0.95906	0.778569	0.9636437	0.930660698	0.922911957	0.96939374	0.9905945	0.74707297
Group 13	0.980884	0.95694	0.824331	0.9749236	0.939074741	0.905769025	0.98751346	0.9779351	0.79450371
Group 15	0.974332	0.95179	0.805496	0.9720985	0.938478071	0.940162479	0.97703673	0.9762925	0.77683967
Group 29	0.965813	0.95353	0.738046	0.9642547	0.924986065	0.933294179	0.96846536	0.9852122	0.70696782
Group 28	0.973221	0.95569	0.840483	0.9701262	0.923798768	0.910928554	0.97652666	0.9920089	0.83713202
Group 34	0.986167	0.96657	0.729892	0.9809594	0.954681358	0.949533779	0.99168573	0.9809756	0.69107638
Group 36	0.9686	0.95365	0.694521	0.9641021	0.931583098	0.918413366	0.9733551	0.9802688	0.63393698

5. Conclusion:

It was found that there is a lot of noise in the data. Such user data had to be discarded. With 10 good user directories, we observe that neural networks performs the best and SVM the worst for user dependent data.