CS 6762

SP, ML, FC

Smart Watch Programming Assignment 2

Group: Alex Fetea, Divya Bagla

Part 1

Previous data had instances. We have 1459 for this assignment after collection.

```
Correctly Classified Instances        1121               76.8334 %
Incorrectly Classified Instances       338               23.1666 %
Kappa statistic                          0.5304
Mean absolute error                      0.2723
Root mean squared error                  0.4035
Relative absolute error                 55.5138 %
Root relative squared error             81.4778 %
Total Number of Instances             1459

=== Detailed Accuracy By Class ===
=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.778    0.245    0.808      0.778    0.793      0.531    0.837     0.844     hand_wash
                 0.755    0.222    0.721      0.755    0.738      0.531    0.837     0.790     no_hand_wash
Weighted Avg.    0.768    0.235    0.770      0.768    0.769      0.531    0.837     0.821

=== Confusion Matrix ===
=== Confusion Matrix ===

   a   b   <-- classified as
 646 184 |   a = hand_wash
 154 475 |   b = no_hand_wash
```

Figure 1 Accuracy result for part 1

The accuracy decreases to 76.83% comparing with 80.27% of assignment 1.

Part 2

Yes, accuracy improves as we increase the time slice.

```
=== Stratified cross-validation ===

Correctly Classified Instances        1107               77.8481 %
Incorrectly Classified Instances       315               22.1519 %
Kappa statistic                          0.5595
Mean absolute error                      0.2579
Root mean squared error                  0.3953
Relative absolute error                 52.3345 %
Root relative squared error             79.6308 %
Total Number of Instances             1422


=== Detailed Accuracy By Class ===
=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.727    0.157    0.855      0.727    0.786      0.567    0.843     0.842     hand_wash
                 0.843    0.273    0.709      0.843    0.770      0.567    0.843     0.771     no_hand_wash
Weighted Avg.    0.778    0.208    0.791      0.778    0.779      0.567    0.843     0.811


=== Confusion Matrix ===
=== Confusion Matrix ===

   a   b   <-- classified as
 579 217 |   a = hand_wash
  98 528 |   b = no_hand_wash
```

Figure 2 Time slice =2

```
=== Stratified cross-validation ===

Correctly Classified Instances        1101               79.4946 %
Incorrectly Classified Instances       284               20.5054 %
Kappa statistic                          0.5922
Mean absolute error                      0.2468
Root mean squared error                  0.372
Relative absolute error                 49.8519 %
Root relative squared error             74.7794 %
Total Number of Instances             1385


=== Detailed Accuracy By Class ===
=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.743    0.141    0.865      0.743    0.799      0.599    0.874     0.865     hand_wash
                 0.859    0.257    0.732      0.859    0.790      0.599    0.874     0.847     no_hand_wash
Weighted Avg.    0.795    0.193    0.805      0.795    0.795      0.599    0.874     0.857


=== Confusion Matrix ===
=== Confusion Matrix ===

   a   b   <-- classified as
 566 196 |   a = hand_wash
  88 535 |   b = no_hand_wash
```

Figure 3 Time slice =3

```
=== Stratified cross-validation ===

Correctly Classified Instances        1112               82.4926 %
Incorrectly Classified Instances       236               17.5074 %
Kappa statistic                          0.6525
Mean absolute error                      0.1987
Root mean squared error                  0.3552
Relative absolute error                 40.0012 %
Root relative squared error             71.2647 %
Total Number of Instances             1348


=== Detailed Accuracy By Class ===
=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.757    0.095    0.903      0.757    0.824      0.663    0.890     0.886     hand_wash
                 0.905    0.243    0.760      0.905    0.826      0.663    0.890     0.852     no_hand_wash
Weighted Avg.    0.825    0.163    0.837      0.825    0.825      0.663    0.890     0.870


=== Confusion Matrix ===
=== Confusion Matrix ===

   a   b   <-- classified as
 551 177 |   a = hand_wash
  59 561 |   b = no_hand_wash
```

Figure 4 Time slice = 4

Part 3

The best time slice from part 2 is 4. The accuracy improves over part 1 but decreases in comparison with part 2, time slice=4.

```
=== Stratified cross-validation ===

Correctly Classified Instances        1067               79.1543 %
Incorrectly Classified Instances       281               20.8457 %
Kappa statistic                          0.5829
Mean absolute error                      0.2447
Root mean squared error                  0.3966
Relative absolute error                 49.2529 %
Root relative squared error             79.5681 %
Total Number of Instances             1348


=== Detailed Accuracy By Class ===
=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.772    0.185    0.830      0.772    0.800      0.585    0.855     0.853     hand_wash
                 0.815    0.228    0.753      0.815    0.782      0.585    0.855     0.800     no_hand_wash
Weighted Avg.    0.792    0.205    0.794      0.792    0.792      0.585    0.855     0.829


=== Confusion Matrix ===
=== Confusion Matrix ===

   a   b   <-- classified as
 562 166 |   a = hand_wash
 115 505 |   b = no_hand_wash
```

Figure 5 Accuracy for part 3

Part 4

The best time interval is 4.

We apply Sequential Feature Selection method and decision tree classifier.

Table 1 Feature names and respective index

Mean X	Std X	Median X	RMS X	Mean Y	Std Y	Median Y	RMS Y	Mean Z	Std Z	Median Z	RMS Z
0	1	2	3	4	5	6	7	8	9	10	11

The feature set is {8, 0, 6, 2}. The accuracy is 83.90% which is an improvement over part 3. The table shows how accuracy improves with feature selection. We stop and don't add the next feature as the improvement is less than 1%.

Table 2 Change in accuracy with each iteration of SFS for DT

Iteration	Feature set	Accuracy (%)
1	{8}	73.664
2	{8, 0}	78.264
3	{8, 0, 6}	81.157
4	**{8, 0, 6, 2}**	**83.902**
5	{8, 0, 6, 2, 10}	84.569

Part 5

When classifier: Random Forest

Table 3 Change in accuracy with each iteration of SFS for Random Forest

Iteration	Feature set	Accuracy (%)
1	{2}	68.249
2	{2, 10}	81.750
3	{2, 10, 6}	86.646
4	**{**2, 10, 6, 1**}**	**88.724**
5	{2, 10, 6, 1, 9}	89.614

Classifier: SVM (weka.classifiers.functions.SMO)

Table 4 Change in accuracy with each iteration of SFS for SVM

Iteration	Feature set	Accuracy (%)
1	{8}	70.919
2	**{8, 5}**	**73.071**
3	{8, 5, 0}	73.813

The accuracy improves as a feature is selected. We stop when improvement in accuracy is less than 1%.

Out of the three classifiers, random forest is the best. It also takes the most time to execute.

- Feature File: Placed all 12 features into features.csv.