Description of the experiments and methodology for solving Worksheet 1 and 3

Himanshu Buckchash
hbuckchash@cs.iitr.ac.in

Nov 2021

1 Worksheet_1 Tasks

Topic 1: Download and preprocess the dataset.
Solution: Performed the required tasks in "dataset.py" file.

Topic 2: Read the research paper and compute the COP and COA features.
Solution: Computed all measures/features using COP. Also, computed COA after creating binary images as suggested in "Topic 2". All the details are available in "worksheet_1_task_1_cop_measures.py". The additional details of the method used for finding AP, ML and COP is described in the section 3 on Methodology.

Topic 3(1): Calculate Euclidean distance and correlation coefficient for separate foot COP features for genuine and imposter samples. And show them in matrices.
Solution: This has been shown in file "worksheet_1_task_2_enrollment.py".

Topic 3(2): Evaluation of key performance metrics.
Solution: This has been shown in file "worksheet_1_task_2_enrollment.py". Specifically, function "task_1_example_of_functionality_for_an_individual()" has shown the calculation of key performance metrics. The results for subject #4, left foot, 0^{th} footstep are shown in Fig. 1 for accuracy plot and in Fig. 2.

Checkpoint: For subject #4 (left foot), using the MDIST features, we had to show minimum distance among templates.
Solution: This has been shown in file "worksheet_1_task_2_enrollment.py". Specifically, function "worksheet_1_checkpoint()" has been used for creating the checkpoint matrices and saving them. They are also available in the "check_point_1_mat" directory.

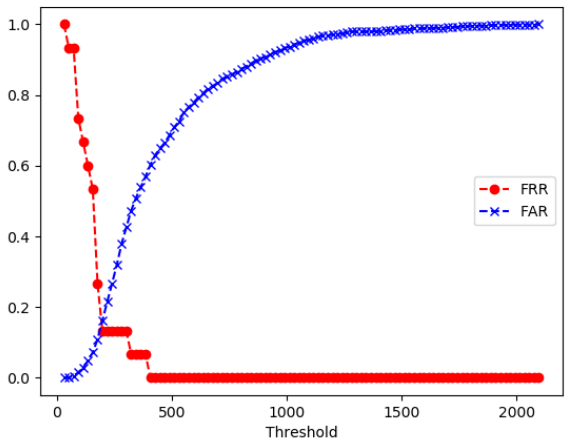

Figure 1: Accuracy plot (FAR, FRR vs threshold) for subject #4, left foot, 0^{th} footstep.

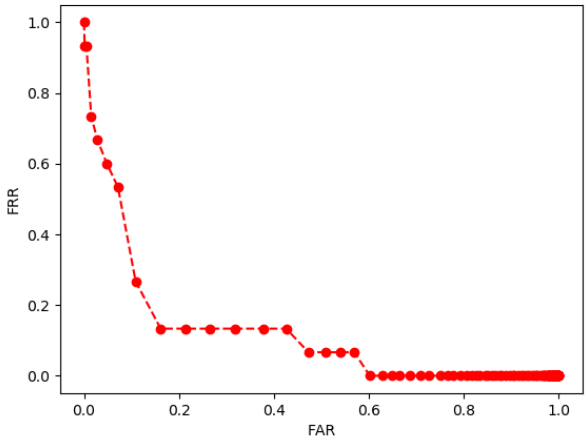

Figure 2: ROC plot (FRR vs FAR) for subject #4, left foot, 0^{th} footstep.

2 Worksheet_3 Tasks

Topic 1: It is explaining the verification stage.
Solution: Read the research paper and the verification stage details.

Topic 2: Feature projection using PCA. Evaluation of multi-template approach and machine learning classifiers. Evaluation of models for using the key performance metrics.
Solution:

- Implemented the standard normalization and PCA.
- PCA was implemented with 19 components and 99.07% explained variance.
- Results of classifiers were less than 30% accuracy scores. All classifiers (linear SVM, LDA, MLP) were used under a one-vs-rest setting and in a classification mode as well. The implementation details can be found in "worksheet_2_task_1.py" file. Classification was performed on 50:50% train/test split of data.
- Results of multi-template approach are shown in file "worksheet_2_task_2_multitemplate.py". In this I have performed the experiment as asked in the worksheet and also as described in the corresponding research paper however I was not able to understand how to implement the FAR and FRR in case of multi-subject evaluation, where we had to choose an average threshold value per subject rather than choosing threshold value per sample of the subject. For this I achieved True Positive Rate (TPR) = 59.55% and False Negative Rate (FNR) = 40.44.

3 Methodology

The literature that I studied is summarized in this google doc. For a given footstep with 100 samples with data shape 60x40x100, 60x40 was considered the size of pressure matrix and 100 as the sampling rate. The pixel positions were considered as the (x,y) coordinates. AP, ML, and COP, for a 40x60 pressure plate matrix, were calculated using equation Eq. 1, 2, 3. In these equations, $P(x, y)$ means pressure at (x, y) coordinates in pressure matrix data. We also calculated the COA after binarizing the pressure matrix using Topic(2,2,c) point. All other COP measures were straight forward to compute by referring to the base research paper mentioned in Worksheet_1.

$$AP = \sum_{x=1}^{40} \sum_{y=1}^{60} \frac{y * P(y, x)}{P(y, x)} \tag{1}$$

$$ML = \sum_{x=1}^{40} \sum_{y=1}^{60} \frac{x * P(y, x)}{P(y, x)} \tag{2}$$

$$COP = (AP, ML) \tag{3}$$

4 My learning from this assignment

A biometric system has two main tasks: (1) verification, (2) identification. Identification is a harder task and is just like any other ML based multi-class classification problem. This assignment was focused on the verification task. Verification involves two stages: (1) enrollment, (2) verification. Enrollment is the most important stage, since it requires the system to find thresholds for each subject based on multiple templates. Selection of the right threshold is important for the appropriate implementation of the biometric system according to the context/application. The threshold is selected as a trade-off between FAR and FRR values. In general, we can select it to minimize the sum of FAR and FRR.

5 Things I could not figure out

Question 1: What is zero-FAR?

6 Improvements or Future works

Deep learning approach: According to my experience, I believe that we can leverage deep learning methods to learn stronger representations for this problem. We may work by:

1. Either deriving primitive measures like AP, ML, RD, COP. And then processing them with deep learning methods for time-series analysis.
2. Or we can learn 2D/3D representations of the pressure matrix and further learn the distribution of these representations in the form of embeddings to leverage deep learning methods for time-series analysis.

Data analysis: I need to analyze the data in order to find the real reasons behind the under-performance of the models, for this I will do following:

1. Visualize the line plots and heat maps of pressure matrices, COP'S and COA's of genuine and impostor_random subject foot.
2. 2d t-SNE Visualization of the COP, AP, ML, RD series' for understanding the variability in the data.
3. Histogram analysis of the measures and AP, ML, COP series.