

Biometric Security (IT499)

Lab Assignment 1

Biometric System Errors and Metrics

Note: Use Python libraries such as numpy, pandas, matplotlib, and sklearn for data handling and plotting, roc_curve and auc functions from sklearn.metrics.

Part 1: Verification Framework

1. Load all four CSV files into your Python environment.
2. Plot the distributions of genuine and impostor scores for both similarity and distance-based systems. (Both genuine and impostor distributions should be probability mass functions, in other words, normalize the frequencies). Plot a decision threshold of 0.5, highlight the portions of the curve leading to errors – False Non-Match Rate (FNMR) and False Match Rate (FMR).
3. Calculate the FNMR and FMR for various threshold values (You may consider a step size of 0.01) and plot FNMR and FMR as a function of the threshold.
4. Determine the threshold at which FNMR equals FMR. Report the EER value for both similarity and distance-based systems. Which biometric system is best based on EER? What is the operating threshold for the systems (decision threshold at EER)?
5. Plot the DET curves (FRR vs. FAR) for all the biometric systems in a single figure for comparison. Plot FAR and FRR on following scale:
 - a. Linear
 - b. normal-deviate
 - c. logarithmic
6. Compare the biometric systems by plotting the Receiver Operating Characteristic (ROC) curves for both similarity and distance-based systems. Use log scale for FAR. Which biometric system is the best one based on ROC curves? Plot on a single figure. Include legend.
7. Calculate the Area Under the ROC Curve (AUC) for each biometric system.
8. Calculate the d-prime and F-ratio values for all systems and comment on the performance of the biometric system.
 - a. Plot the genuine and impostor distributions of the original system,
9. In the same figure, plot the genuine and impostor distributions after removing the goats and lambs. Consider top 5% scores with highest FNMR (goats) and top 5% of Consider Doddington's classification for one of the systems and answer the following:
 - a. scores with highest FMR (lambs). Plot the mean and standard deviation of their genuine scores and impostor scores in all three cases. Use legend to differentiate between the three cases.

The following four CSV files are provided for this assignment corresponding to the four biometric systems:

1. System 1: similarity_scores1.csv
2. System 2: similarity_scores2.csv

3. System 3: distance_scores1.csv
4. System 4: distance_scores2.csv

Each file contains the following columns:

- `score`: The comparison score (either similarity or distance-based).
- `label`: Indicates whether the score is genuine (`g`) or impostor (`i`).

-
- Submit your Jupyter/Colab notebook with the complete code and outputs.
 - Naming convention: Rollno_Labxx.ipynb
 - Include a written report summarizing your observations, outputs and conclusions in a separate pdf file.