TWO TYPES OF FACE RECOGNITION AND METRICS FOR EVALUATION

By Li Peifeng

TYPES OF FACE RECOGNITION

There are 2 types of human face recognition: verification and identification. Verification is a task to determine whether 2 input images are about a same person, so it's also called 1:1 matching. Identification is another task to find out which image in the gallery contains the person most similar to the one in an input image, so it's called 1:N matching. Identification task can be further divided into close-set (the person in the input image is included in the gallery for certain) and open-set (the person in the input image is not necessarily included in the gallery). For easier understanding, verification answers 'are you who you are', and identification answers 'who you are', so the metrics for these 2 types of tasks are slightly different.

METRICS FOR FACE VERIFICATION

For face verification, there are only 4 types of results (you are the person, you aren't the person, you are the person but actually you aren't, you aren't the person but actually you are), so the way to evaluate performance is similar to the confusion matrix used in binary classification, but there are some slight differences. Let's look at the confusion matrix first:

		Actual Values: Positive (1)	Actual Values: Negative (0)
Predicted Values	Positive (1)	TP	FP
Predicted Values	Negative (0)	FN	TN

Still, $Acc = \frac{TP+TN}{All}$. Besides, we have:

1. $TAR = \frac{TP}{TP+FN} = Recall$. TAR stands for True Accept Rate, indicating the rate of persons who should be accepted are finally accepted by the model.

2. $FAR = \frac{FP}{FP+TN}$. FAR stands for False Accept Rate, indicating the rate of persons who should be rejected are finally accepted by the model. In practice, it means letting an unauthorized person in.

3. $FRR = \frac{FN}{TP+FN} = 1 - TAR$. FRR stands for False Reject Rate, indicating the rate of persons who should be accepted are finally rejected by the model. In practice, it means rejecting an authorized person to get in.

METRICS FOR FACE IDENTIFICATION: IN CLOSE-SET

In a close-set identification task, Rank-K @ gallery size is used for evaluation. For this type of identification, a probe (test image) is necessarily included in the gallery. In a test, calculations of similarities between the probe and all gallery images are first performed, after which gallery images will be ranked by similarity (from high to low). A top-K accuracy then will be calculated (if a correct matching is in top-K results, then a hit is counted). After all probes are tested, Rank-K @ gallery size can be represented as:

$$\frac{\#hits}{\#probes} @ \#gallery\ images$$

METRICS FOR FACE IDENTIFICATION: IN OPEN-SET

Different to close-set, a probe is unnecessarily included in the gallery. For this reason, open-set identification also needs to determine whether the probe is in the gallery, which makes metrics like a combination of verification and close-set identification. First, let's consider 5 situations in an open-set identification:

1. A probe is matched with a gallery image, and the probe is about the person in that image, which is called IBC (a probe **in** gallery and its similarity with a certain gallery image is **bigger** than the threshold for distinguishing whether a probe is about a person in gallery, and it's matched with the **correct** person).
2. A probe is matched with a gallery image, but the probe is about another person in the gallery, which is called IBE (in & bigger & error).
3. A probe isn't matched with any gallery image, but the probe is about a person in the gallery, which is called IS (in & smaller).
4. A probe is matched with a gallery image, but the probe is about a person **not in the gallery**, which is called OB (out & bigger).
5. A probe isn't matched with any gallery image, and the probe is about a person not in the gallery, which is called OS (out & smaller).

Then we can have these metrics for evaluation:

1. $Rank - K@FAR = \frac{\#IBC_k + \#OS_k}{\#probes} @ \frac{\#OB}{\#probes\ not\ in\ gallery}$. It's a higher-better metric. IBC_k means only using top-K results of a probe as the gallery to determine whether it's and IBC. (same to OS_k, so '0' might be counted if a probe doesn't meet definitions of IBC and OS when the gallery is limited to its top-K results). Generally, Rank-K is similar to Accuracy in verification, and FAR indicates the rate of persons who should be rejected are finally accepted by the model (same as verification).

2. $TAR@FAR = \frac{\#IBC}{\#probes\ in\ gallery} @ \frac{\#OB}{\#probes\ not\ in\ gallery}$. It's also a higher-better metric. TAR is slightly different to the TAR in verification, for this TAR only counts persons who should be accepted are finally **correctly accepted** (not accepted as another authorized person).

3. $FRR@FAR = \frac{\#IS}{\#probes\ in\ gallery} @ \frac{\#OB}{\#probes\ not\ in\ gallery}$. It's a lower-better metric. FRR still indicates the rate of persons who should be accepted are finally rejected (same as verification).