.cse@buffalo

Predicting Performance in Large-Scale Identification Systems by Score Resampling

Sergey Tulyakov, Venu Govindaraju

Center for Unified Biometrics and Sensors,
University at Buffalo

Problem Statement

cse@buffalo

Test system

Number of users: $G_T = 100$

$$x(1) \quad y_1(1) \quad y_2(1) \quad \ldots \quad y_{G_T}(1)$$

$$x(2) \quad y_1(2) \quad y_2(2) \quad \ldots \quad y_{G_T}(2)$$

$$\ldots$$

i-th test identification trial

$$x(i) \quad y_1(i) \quad y_2(i) \quad \ldots \quad y_{G_T}(i)$$

$$\ldots$$

Genuine Scores **Impostor Scores**

Predicted system

$$G = 3000$$

$$x(j) \quad y_1(j) \quad y_2(j) \quad \ldots \quad y_G(j)$$

Need to estimate correct identification rate:

$$CIR = P(x(j) > \max_{1 \le k \le G} y_k(j)) = ?$$

- Predicting performance in *closed set identification systems*
- NIST BSSR1 score set used in experiments

Analysis of previous approaches

- Score Mixing Effect

Action:

$$x(i) \quad y_1(i) \quad y_2(i) \quad \dots \quad y_{G_T}(i)$$

$$\dots$$

$$x(k) \quad y_1(k) \quad y_2(k) \quad \dots \quad y_{G_T}(k)$$

Mix scores from different test identification trials during prediction
(e.g. to estimate impostor score density)

Effect:

- assumption of iid genuine and impostor score from different trials (usually not true)
- underestimation of CIR

- Binomial Approximation Effect

Action: try to approximate

$$CIR = \int_{-\infty}^{+\infty} N^G(t) m(t) dt$$

cumulative distribution function of impostor scores

density of genuine scores

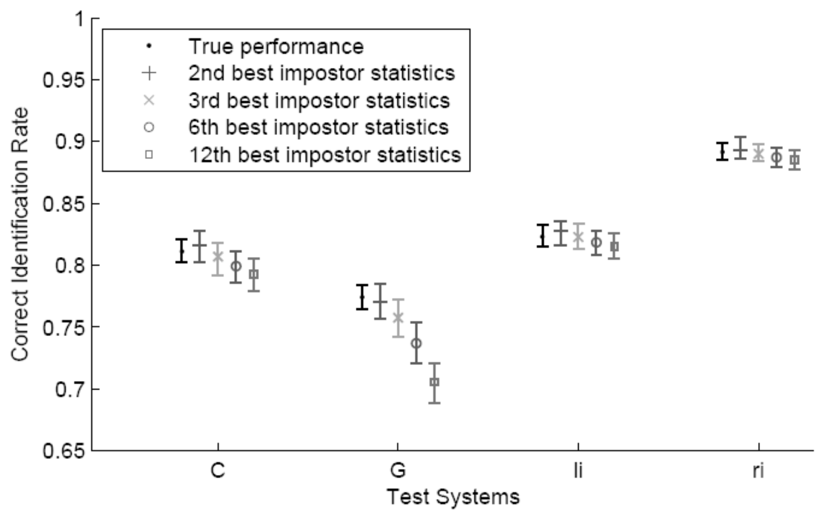
Effect:

- small errors in estimating $N(t)$ lead to large errors in $N^G(t)$
- overestimation of CIR

Proposed approach

- <u>Score Resampling</u>
  - basic simulation of larger identification system workflow
  - genuine and impostor scores are chosen from available test scores

- <u>No binomial approximation effect</u>
- <u>Able to control score mixing effect</u>
  - significantly less score mixing is needed than for binomial approximations
  - mix scores only from similar test identification trials
  - determine similarity of identification trials by properly chosen statistics
  - nth order statistics seems to be appropriate for most matchers



Summary

- Simple approach for predicting identification system performance
- Lesser requirements on test data than for binomial approximation based methods
- Theoretically justified (larger version of the paper)
- Precise (depending on properly chosen statistics for score mixing)
- Easily extended for predicting performance in open set identification systems